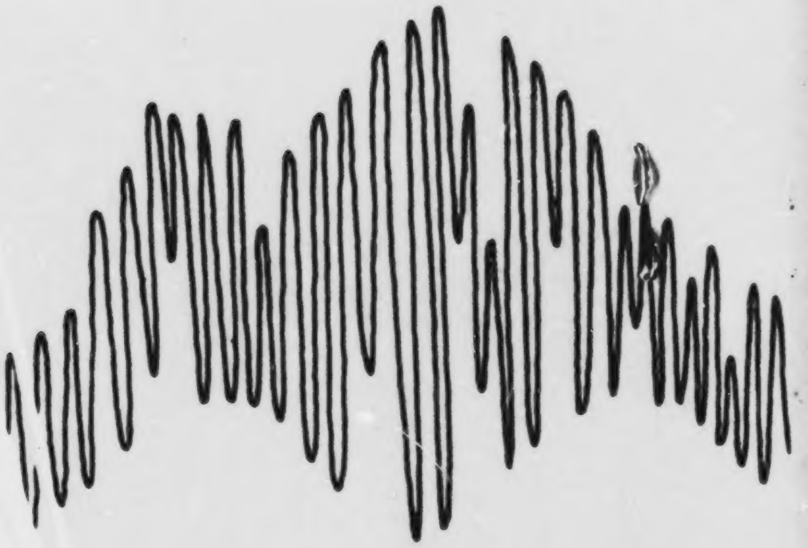
SPECTRAL ANALYSIS

JUL 1970
RECEIVED
NASA STI FACILITY
ACQ BRANCH



APPLICATIONS IN WATER POLLUTION CONTROL

FACILITY FORM 602

N70-35827
(ACCESSION NUMBER) (THRU)
104
(PAGES) (CODE)
3
(NASA CR OR TMX OR AD NUMBER) (CATEGORY)
06

U.S. DEPARTMENT OF THE INTERIOR • FEDERAL WATER POLLUTION CONTROL ADMINISTRATION

SOT

SPECTRAL ANALYSIS

APPLICATIONS IN WATER POLLUTION CONTROL

by

T. A. WASTLER

Chief, Engineering and Physical Sciences Section
Estuarine and Oceanographic Programs Branch
Division of Planning and Interagency Programs
Office of Operations
Federal Water Pollution Control Administration
U.S. Department of the Interior
Washington, D.C. 20242

December 1969

CWT-3

PRECEDING PAGE BLANK NOT FILMED.

CONTENTS

Page

Preface -- vi

Chapter 1. The Concept of Spectral Analysis ---------------- 1
Chapter 2. The Techniques of Spectral Analysis ------------- 7
Chapter 3. Interpretation of Spectra ----------------------- 19
Chapter 4. Design Criteria for Spectral Analysis ----------- 27
Chapter 5. Introduction to Cross-Spectra ------------------- 31
Chapter 6. Computation of Cross-Spectra -------------------- 35
Chapter 7. Interpretation of Cross-Spectra ----------------- 41
Chapter 8. Spectral Analysis as a Data Processing System ------ 47

Example 1. Tracing a Conservative Pollutant in a River 49

Example 2. Dilution and Decay of a Nonconservative Pollutant in a River ---------------- 59

Example 3. Diffusion of a Nonconservative Pollutant in an Unstratified Tidal River --------- 73

Example 4. Relationship of Change in River Flow to Temperature and Dissolved Oxygen in an Unstratified Tidal System --------- 84

Example 5. Relationship of River Flow and Tide to Stratification in an Estuary ----------- 94

Acknowledgement -- 98

Bibliography --- 99

TABLES

Table 1.a. Input Record. Salt Waste Discharge 50
Table 1.b. Output Record. Water Supply Hardness 51
Table 2. Recurrence Periods at Various Choices of Total Lags 53
Table 3. Computer Output of Cross-Spectral Analysis: Salt Waste—Water Hardness Problem 54
Table 4. Comparison of Salt Hardness Values 60
Table 5. Correlation Coefficients of Response Spectra to Outfall Discharges 69
Table 6. Phase Lags and Responses to Outfall Discharges 70
Table 7. Times of Passage Calculated from Diurnal Response Components 71
Table 8. Dispersion of Wastes Between River Stations 72
Table 9.a. Individual Power Spectra of Water Temperature 88
Table 9.b. Water Temperature Response to River Discharge 89
Table 9.c. Stream Dissolved Oxygen Response to Temperature 90
Table 9.d. Stream Dissolved Oxygen Response to River Discharge 91
Table 10. Dissolved Oxygen, Temperature, and River Discharge Relationships 92
Table 11. Overall Response of Chloride Concentration to River Discharge, Charleston Harbor 96
Table 12. Surface- to Bottom-Chloride Ratios, Charleston Harbor 97
Table 13. Analysis of Chloride Responses to Tide Ranges, Charleston Harbor 98

FIGURES

Figure 1. Physical Analogy to Spectral Analysis 2
Figure 2. Typical Spectra Obtained from Several Types of Curves 4
Figure 3. Portion of Tide Height Record in the Potomac Estuary 8
Figure 4. Spectrum of a Tidal Height Record 12
Figure 5. Spectrum of a Tidal Height Hecord 14
Figure 6. Spectrum of a Tidal Height Record 15
Figure 7. Dissolved Oxygen Records Obtained in the Potomac Estuary, August 1959 20
Figure 8. Biochemical Oxygen Demand Records Obtained in the Potomac Estuary, August 1959 21
Figure 9. Dissolved Oxygen Spectra in the Potomac River 22
Figure 10. Biochemical Oxygen Demand Spectra in the Potomac Estuary 23
Figure 11. Schematic Representation of Potomac Estuary Sampling Station Locations 24

Figure 12. Typical Input-Output System ---------------------- 33
Figure 13. Geometrical Analogy to Cross-Spectral Analysis ----- 38
Figure 14. Comparison of Typical Data Records -------------- 43
Figure 15. Dissolved Oxygen Variance as a Function of Mean Dissolved Oxygen, Potomac Estuary, August 1959 -- 45
Figure 16. Factors Affecting Stream Dissolved Oxygen Concentration ------------------------------- 48
Figure 17. Spectra from Salt Hardness Analysis -------------- 55
Figure 18. Schematic Diagram Showing Locations of Outfalls and Sampling Stations ------------------------ 61
Figure 19. Waste Discharge from Outfall A, Individual Samples -- 62
Figure 20. Waste Discharge from Outfall B, Individual Samples -- 63
Figure 21. Spectra of Waste Discharge Parameters, Outfall A ---- 64
Figure 22. Spectra of Waste Discharge Parameters, Outfall B ---- 65
Figure 23. Spectra of Stream Biochemical Oxygen Demand with Biochemical Oxygen Demand Response Spectra --- 67
Figure 24. Spectrum of Solar Radiation --------------------- 74
Figure 25. Stream Dissolved Oxygen Spectra in the Potomac Estuary ------------------------------------ 75
Figure 26. Stream Biochemical Oxygen Demand Spectra in the Potomac Estuary ---------------------------- 76
Figure 27. Stream Dissolved Oxygen Response to Sunlight ------ 77
Figure 28. Relationship of Biochemical Oxygen Demand to Dissolved Oxygen Response to Solar Radiation ----- 78
Figure 29. Corrected Stream Biochemical Oxygen Demand Spectra ----------------------------------- 80
Figure 30. Dissolved Oxygen Response to Corrected Stream Biochemical Oxygen Demand ------------------ 81
Figure 31. Volumes of Water Present in the Potomac Estuary at Washington, D. C. ------------------------ 82
Figure 32. Dissolved Oxygen Response to Temperature --------- 85
Figure 33. Spectra of Driving Forces and Water Temperature ---- 86
Figure 34. Charleston Harbor Sampling Locations ------------ 95

PREFACE

This volume is designed to offend almost everyone engaged in the control of water pollution. Sanitary engineers will be offended because there is too much mathematics; mathematicians, statisticians, and systems analysts will object because the mathematics is over-simplified; administrators and managers will point out that there is no panacea here for management decisions; oceanographers will be shocked because the author has felt it necessary to explain a technique which has been common practice for many years in oceanography.

All of these criticisms are just. Spectral analysis has been used for many years in oceanography, meteorology, electrical engineering, medicine, and several other fields. It is a method of data analysis and, as such, is a contributor toward the technical foundation upon which management can make a decision, but it provides no magic mathematical model to assist in the process. The mathematics and statistics of what is essentially a mathematical and statistical tool are indeed over-simplified; but the engineering rationale is also over-simplified to make it comprehensible to mathematicians and systems analysts.

In short, the author has attempted the dangerous task of bridging fields and trying to explain one field to another. The concepts and techniques of computation presented here are merely the author's interpretation of what others have developed—for these ideas he takes no credit. The interpretations of spectral results in terms of sanitary engineering phenomena are entirely the author's ideas—for these he takes all the blame.

The past decade has seen great changes in the basic approach to the study of water pollution and water resources problems. The advent of high-speed digital computers has resulted in a spate of mathematical models based on the use of the rather sparse data obtained by manual techniques. At the same time advances in instrumentation, recording, and telemetry have made remote sensing stations operational in environmental monitoring, but the use of data from them has been minimal in water pollution control.

It is the gap between environmental monitoring and mathematical modeling that the author has attempted to bridge by pointing out the application of spectral analysis in the interpretation of time series data, both in estuaries and in free-flowing streams.

Spectral analysis as a technique emerged from the work of La Place and Fourier near the beginning of the nineteenth century, and much of the mathematical theory underlying its practical application was developed in the work of G. I. Taylor, Norbert Wiener, and S. O. Rice. With the advent of high-speed digital computers and the development of automatic sampling techniques its practical use in many fields became feasible. The author offers his own experiences in the use of spectral analysis in water pollution problems as a first attempt in the use of this technique in the study of water pollution in rivers and estuaries.

T. A. WASTLER
WASHINGTON, D. C.
DECEMBER 1968

1

THE CONCEPT OF SPECTRAL ANALYSIS

One of the classic experiments in the study of optics is to pass a beam of sunlight through a triangular prism and observe the rainbow of colors into which the beam of white light is split. This experiment serves as a useful physical analogy to the operation of spectral analysis on a body of data.

In Figure 1a is presented a schematic diagram of the results of passing a beam of sunlight through a triangular prism. The beam of sunlight is split into a spectrum of colors ordered according to their respective wave lengths (or frequencies). If this spectrum of colors is allowed to strike a battery of light-sensitive cells, the intensity of the light of each color can be measured; these results can then be plotted as a "light intensity spectrum," which might look like that in Figure 1a if each of the six principal colors had the same intensity.

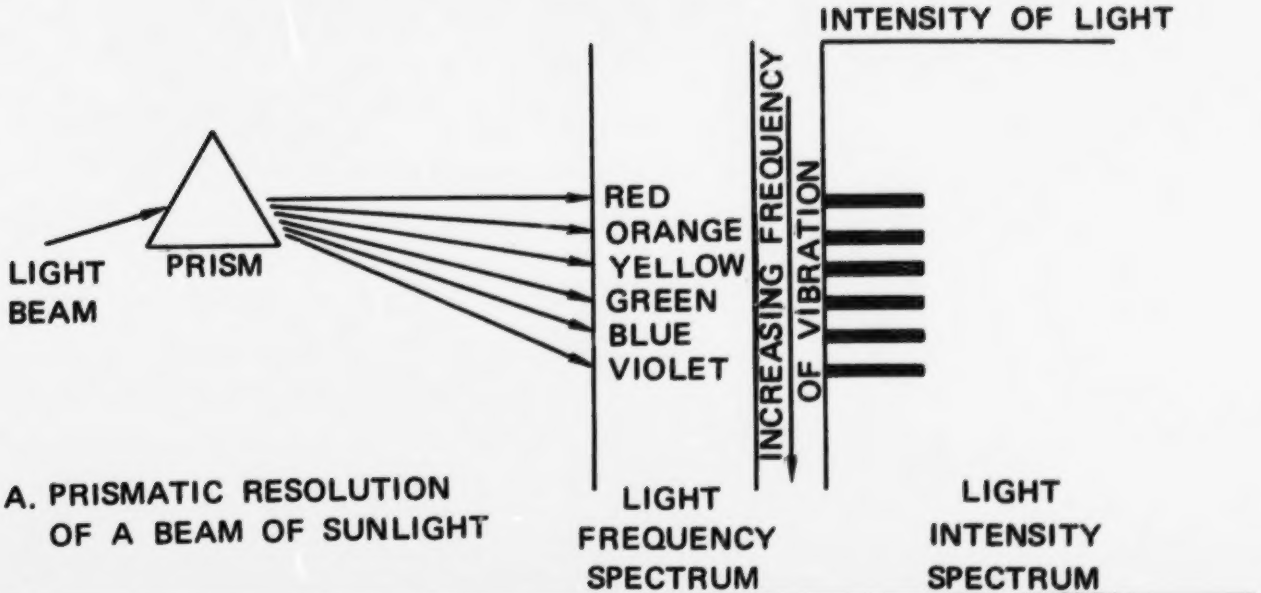
If a similar experiment is performed with a light beam that is composed of only three of these colors (present at different intensities), the resulting light frequency spectrum and light intensity spectrum might resemble those in Figure 1b.

This experiment demonstrates the resolution of a complex physical phenomenon into a group of simpler phenomena that may be easier to examine from both theoretical and practical viewpoints.

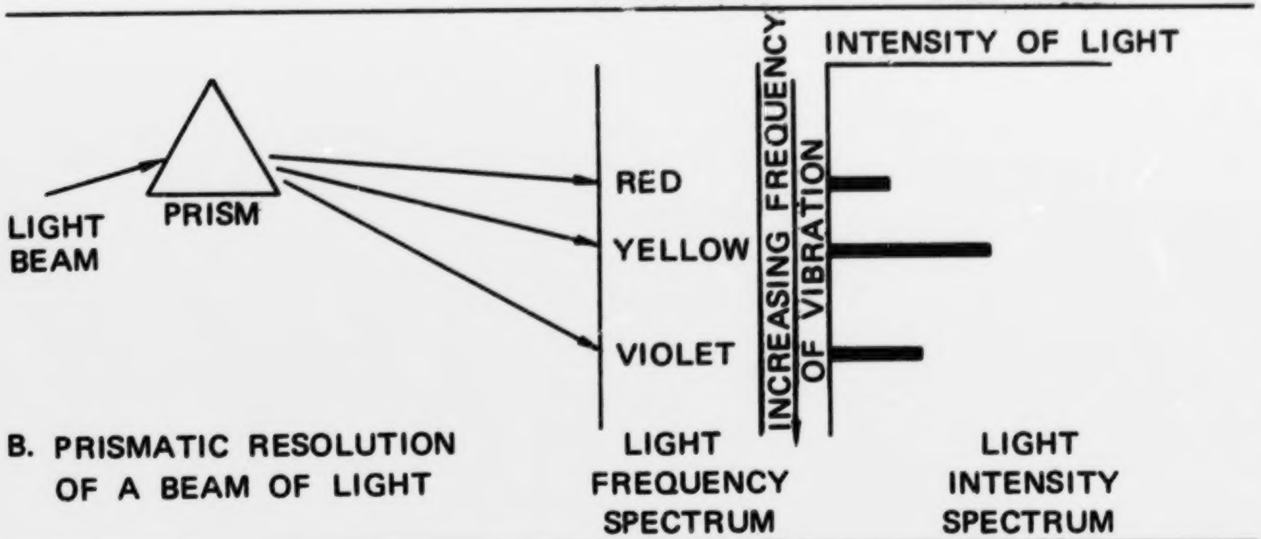
The effect of using spectral analysis on a record of observed field data is directly analogous to the effect of the prism on the light beam. This analogous effect is presented schematically in Figure 1c. The actual technique of spectral analysis is discussed later; at this time the significance of the result of the computation is of concern.

In the prism experiment the relative intensities of the resolved light frequencies can be observed and studied. Spectral analysis of a record of observations results in a sorting of the total "variance" of the record into its component frequencies. The variance of a data record is therefore analogous to the intensity of the light beam.

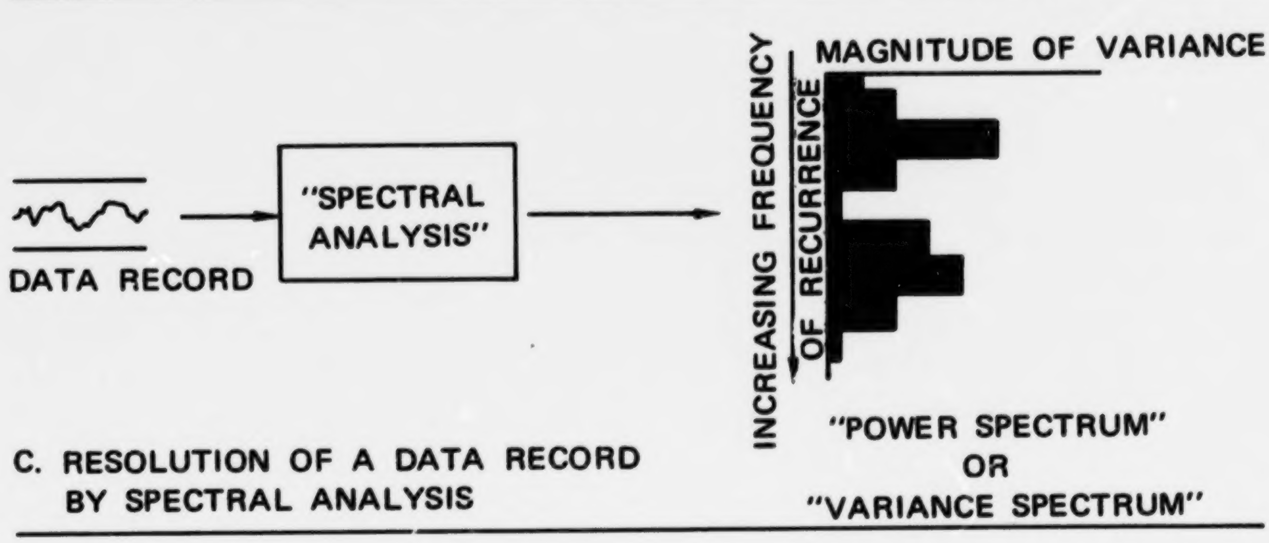
The *variance* is defined as the sum of the squares of the deviations from the mean divided by one less than the number of observations. This is the definition of variance as it is normally used as a descriptive statistic of a


INTENSITY OF LIGHT
LIGHT BEAM
PRISM
RED
ORANGE
YELLOW
GREEN
BLUE
VIOLET
INCREASING FREQUENCY OF VIBRATION
LIGHT FREQUENCY SPECTRUM
LIGHT INTENSITY SPECTRUM

A. PRISMATIC RESOLUTION OF A BEAM OF SUNLIGHT


INTENSITY OF LIGHT
LIGHT BEAM
PRISM
RED
YELLOW
VIOLET
INCREASING FREQUENCY OF VIBRATION
LIGHT FREQUENCY SPECTRUM
LIGHT INTENSITY SPECTRUM

B. PRISMATIC RESOLUTION OF A BEAM OF LIGHT


MAGNITUDE OF VARIANCE
DATA RECORD
"SPECTRAL ANALYSIS"
INCREASING FREQUENCY OF RECURRENCE
"POWER SPECTRUM" OR "VARIANCE SPECTRUM"

C. RESOLUTION OF A DATA RECORD BY SPECTRAL ANALYSIS

Figure 1. Physical Analogy to Spectral Analysis.

body of data. Conceptually, the variance is a measure of the dispersion of observations about the mean value. In the statistical interpretation of data, it is ordinarily regarded that this dispersion of values about the mean is due to chance.

The statistical manipulation of time-series data by spectral analysis results in the computation of those parts of the variance of a record that recur at constant time intervals as well as the part that is random (nonrecurring) in character. As the light beam may be resolved by the prism into its component colors of different intensity, so is the variance of a time-series record resolved into its component parts by spectral analysis.

The estimates of variance for each frequency resolved in the spectral analysis form the "power spectrum" of the record from which the computations are made. (The term "variance spectrum" would be more accurate; but in the pioneering work in spectral techniques done in communications engineering the term "power," which is closely related to record variance in that frame of reference, came into common usage.) The power spectrum computation is the fundamental operation of data reduction and interpretation by means of spectral analysis; the many other computations that can be made in spectral analysis are all based firmly upon the concept and calculation of individual power spectra. It may be stated that the computation of individual power spectra bears about the same relationship to spectral analysis as differentiation and integration bear to the Calculus.

The interpretation of variance as a statistic descriptive of both the random and nonrandom characteristics of a time-series data record is most important in understanding the significance of spectral results. In the usual type of statistical analysis, variance is conceptually regarded as a measure of the random dispersion of the observations from their mean value. In many cases this is true; but that this is not a necessary condition for the existence of a variance can be demonstrated with the aid of Figure 2, in which segments of three hypothetical records and the corresponding power spectra are presented.

Figure 2a shows a record that has a constant value, *i.e.,* all values are equal to the mean. Since there are no deviations from the mean, the variance is zero and the power spectrum is zero at all frequencies.

Figure 2b shows a record that forms a sloping straight line. The segment of the record shown in this figure has a mean of 3.55 and a variance of 0.69. It is apparent that none of this variance results from a "random" dispersion about the mean, but that the variance is the result of a secular (time-dependent) trend in the record. If spectral analysis were done on the record of which this segment is a part, all of the variance (or "power") would be concentrated in the zero-frequency spectral estimate as shown in Figure 2b. The zero-frequency spectral estimate includes all of the record variance that does not recur during the length of the record used in the analysis. It therefore includes (1) any truly random fluctuations in the record, (2) any linear trends in the record, and (3) any periodic components in the record that are of so low a frequency that they appear as linear trends in the record. For example, spectral analysis of the segment "A" in Figure 2c would result in a power spectrum similar to that in Figure 2b, simply because the record length is not great enough to resolve the periodic fluctuation exhibited in Figure 2c.

The record of which Figure 2c is a small part also exhibits a variance. The mean of this record is 4.0, and the variance is entirely the result of the sinusoidal fluctuation about the mean. Spectral analysis of this record, which should be about 10 times longer than the segment presented in Figure 2c, would result in a power spectrum in which all of the variance

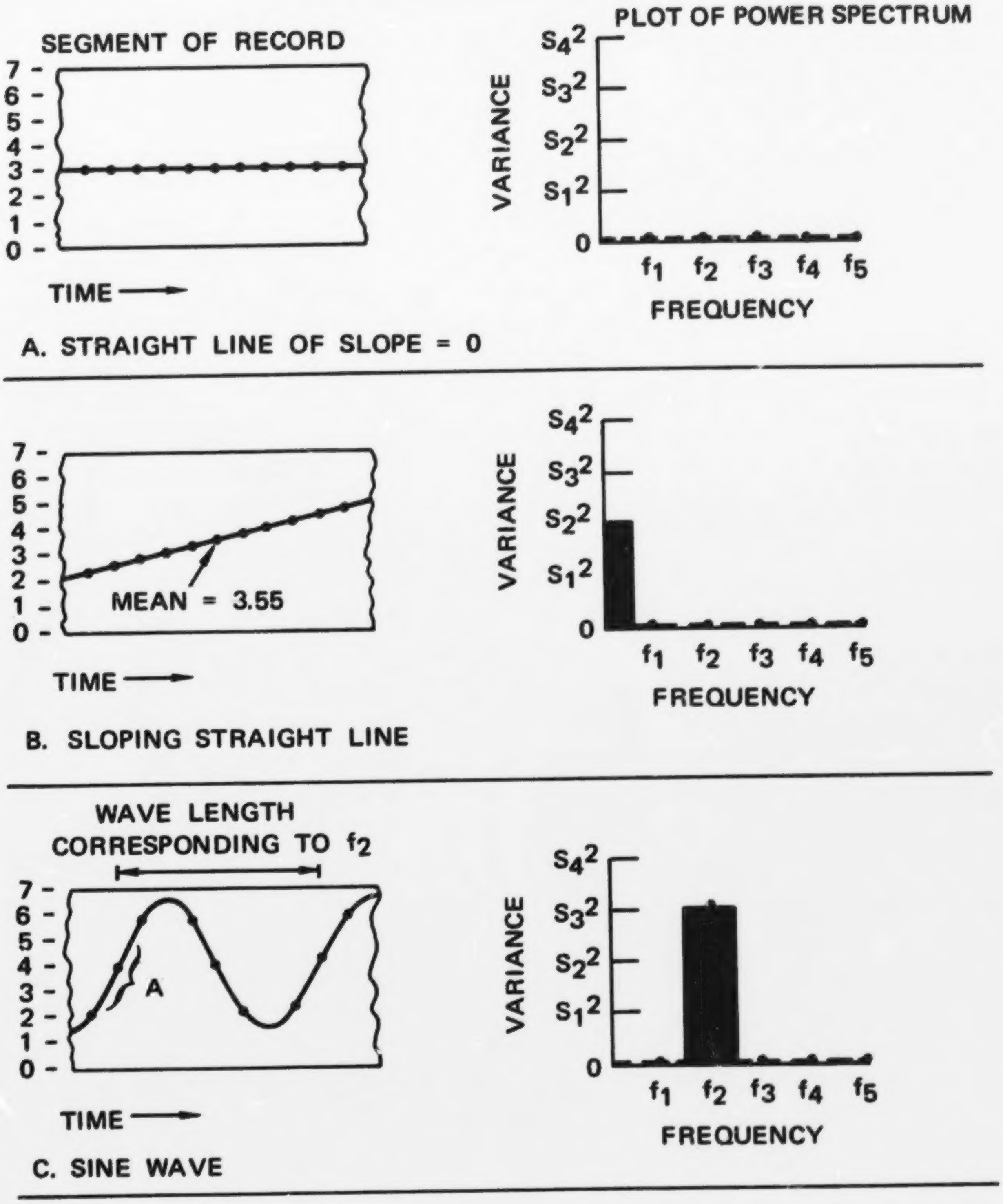


Figure 2. Typical Spectra Obtained from Several Types of Curves.

is concentrated at the frequency designated f_2, which corresponds to the wave length of the sine wave in the data record.

If the data record were a combination of Figures 2b and 2c, the power spectrum would be a combination of the spectra in these figures, *i.e.,* there would be components at the zero frequency and at f_2. It is this characteristic that makes spectral analysis such a useful tool in analyzing records that represent complex phenomena in natural systems. For example, the diurnal effects of photosynthetic activity could be separated from the longer-period effects of waste loads and river discharges by spectral analysis of the stream dissolved oxygen (DO) record.

The power spectra in Figures 1 and 2 are presented in bar graphs to emphasize two characteristics of spectral results: (1) estimates of the variance at several discrete frequencies are obtained from the analysis and (2) each variance estimate represents the variance concentrated in a band around the nominal frequency of each variance. This bar graph representation is not the usual way in which spectra are presented; in the remaining figures discussed in this paper, the conventional point-and-line representation is used.

Like any other method of data analysis, this method has its limitations and disadvantages. Three major requirements for the record may be regarded as limiting. First, the record must be fairly long—generally having over 100 sequential measurements. Second, there must be no missing data—if measurements are missing, suitable values must be interpolated before spectral analysis is attempted. Third, the mathematical procedures require so much computation that the use of a high-speed digital computer is essential for most analyses. These restrictions are discussed in more detail.

PRECEDING PAGE BLANK NOT FILMED.

2

THE TECHNIQUES OF SPECTRAL ANALYSIS

The approach used here is to present the technique of spectral analysis by following through the steps in the actual spectral analysis of a record. For the benefit of those who wish to explore the mathematical basis for spectral analysis computations, pertinent references are presented in the bibliography.

The record chosen for detailed examination here is a record of water level obtained from a U.S. Geological Survey station in the Potomac Estuary near Washington, D. C. This record was chosen because it exhibits a simple periodicity with very little random interference ("noise"), and because the data were obtained as a continuous recording so that a wide choice of sampling intervals was possible.

A portion of this record is presented in Figure 3. A visual examination of this record shows that there is a dominant period of about 12 hours and that there is some long-period change.

Although the entire computation can be carried out on high-speed digital computers, with available programs, the individual steps are presented here to illustrate the technique. Only the initial steps in the data preparation need be carried out by hand or on semi-automatic equipment.

Step 1. A sampling interval of four hours was chosen for this particular analysis, and the record was read at this interval for a total of 145 readings. (Considerations governing the number of points read and the sampling interval are discussed later in detail.) The starting point was arbitrary. The numbers obtained, in order, are

1.30 = value 1
2.57 2
3.79 3
1.49 4
2.30 5
4.73 6
• •
• •
• •

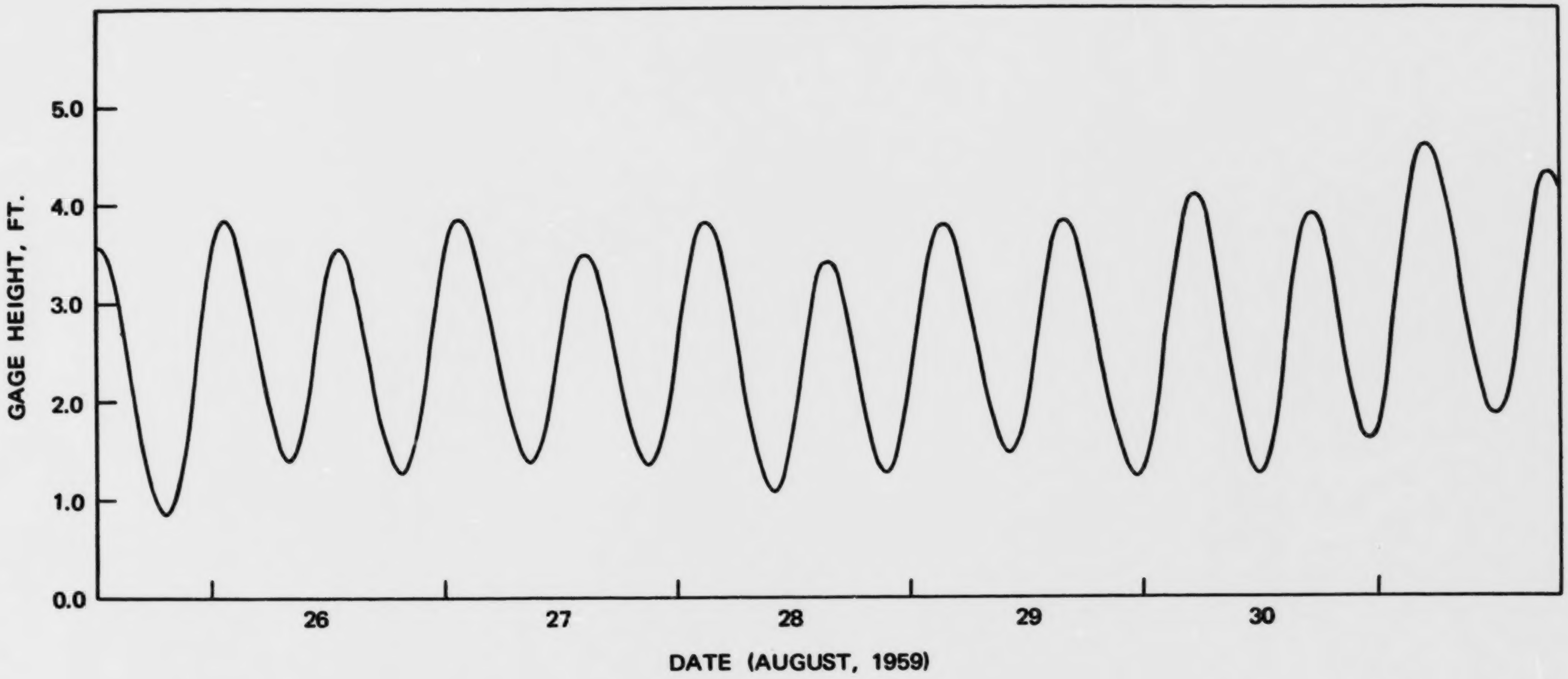


Figure 3. Portion of Tide Height Record in the Potomac Estuary.

•	•
•	•
•	•
3.10	140
1.46	141
3.16	142
3.30	143
1.41	144
2.35	145

Mean = 2.43
Square of Mean = 5.90

Step 2. The autocorrelation function of these numbers is then formed. This is a very large name for a very simple, but very useful, process. Each number in the record is multiplied by another number in the record, and from the mean of the sum is subtracted the square of the arithmetic mean of the entire record. If the individual products are formed by multiplying each number by itself, the result is called the "autocorrelation at lag 0," which is purely and simply the "variance" as ordinarily defined in statistics. If the individual products are formed by multiplying each number by the number that follows it in the sequence, the result is called the "autocorrelation at lag 1." The autocorrelations computed to 12 lags for the record being analyzed here are

$C_0 \equiv$ Autocorrelation at lag 0
1.30 × 1.30 = 1.69
2.57 × 2.57 = 6.60
3.79 × 3.79 = 14.36
1.49 × 1.49 = 2.22
2.30 × 2.30 = 5.29
4.73 × 4.73 = 22.37
•   •
•   •
•   •
•   •
3.10 × 3.10 = 9.61
1.46 × 1.46 = 2.13
3.16 × 3.16 = 9.99
3.30 × 3.30 = 10.89
1.41 × 1.41 = 1.99
2.35 × 2.35 = 5.82

Sum = 1046.9

$$\begin{aligned} \frac{1046.9}{145} &= 7.22 \\ &\underline{-5.90} \\ C_0 &= 1.32 \end{aligned}$$

$C_1 \equiv$ Autocorrelation at lag 1
1.30 × 2.57 = 3.34
2.57 × 3.79 = 9.74
3.79 × 1.49 = 5.65
1.49 × 2.30 = 3.43
2.30 × 4.73 = 10.88
4.73   •   •
•   •   •
•   •   •
•   •   •
•   •   •
3.10 × 1.46 = 4.53
1.46 × 3.16 = 4.61
3.16 × 3.30 = 10.43
3.30 × 1.41 = 4.65
1.41 × 2.35 = 3.31
2.35

Sum = 804.67

$$\begin{aligned} \frac{804.67}{144} &= 5.588 \\ &\underline{-5.90} \\ C_1 &= -0.312 \end{aligned}$$

$C_2 \equiv$ Autocorrelation at lag 2

1.30 × 3.79 = 4.93
2.57 × 1.49 = 3.83
3.79 × 2.30 = 8.72
1.49 × 4.73 = 7.05
2.30 × • •
4.73 × • •
• • •
• • •
• • •
• • •
3.10 × 3.16 = 9.80
1.46 × 3.30 = 4.82
3.16 × 1.41 = 4.46
3.30 × 2.35 = 7.76
1.41
2.35

Sum = 764.62

$$\begin{array}{rcr} \dfrac{764.62}{143} & = & 5.347 \\ & & \underline{-5.90} \\ C_2 & = & -0.553 \end{array}$$

$C_3 \equiv$ Autocorrelation at lag 3

1.30 × 1.49 = 1.94
2.57 × 2.30 = 5.91
3.79 × 4.73 = 17.93
1.49 × • •
2.30 × • •
4.73 × • •
• • •
• • •
• • •
• • •
3.10 × 3.30 = 10.23
1.46 × 1.41 = 2.06
3.16 × 2.35 = 7.43
3.30
1.41
2.35

Sum = 1015.3

$$\begin{array}{rcr} \dfrac{1015.3}{142} & = & 7.10 \\ & & \underline{-5.90} \\ C_3 & = & 1.20 \end{array}$$

The remaining autocorrelations are computed similarly and have the values

$C_4 = 0.154$
$C_5 = 0.735$
$C_6 = 1.10$
$C_7 = 0.0476$
$C_8 = -0.917$
$C_9 = 0.921$
$C_{10} = 0.252$
$C_{11} = -0.998$
$C_{12} = 0.798$

This operation may be expressed mathematically as

$$C_r = \frac{1}{n-r} \sum_{1}^{n-r} x_t x_{t+r} - \left[\frac{1}{n} \sum_{1}^{n} x_t \right]^2, \qquad (1)$$

where

C_r = autocorrelation at lag r,
x_t = record value at t,
t = 0, 1, 2 . . . n = sequential index of values,
r = 0, 1, 2 . . . m = lag number,
m = total number of lags.

Step 3. The Fourier cosine transform for each autocorrelation is next computed. This serves to smooth out some of the wide fluctuations in the autocorrelations and consists of applying a cosinusoidal weighting factor to each autocorrelation calculated in the preceding step. This operation can be expressed mathematically as

$$V_r = \frac{k}{m}\left[C_0 + C_m \cos r\pi + 2\sum_{q=1}^{m-1} C_q \cos\frac{qr\pi}{m} \right], \qquad (2)$$

where

V_r = Fourier cosine transform of the autocorrelation at lag r,
q = lag number, having values between 1 and $m-1$,
k = a constant,

$$k = 1 \text{ for } r = 1, 2 \ldots m-1,$$
$$k = \tfrac{1}{2} \text{ for } r = 0,\ r = m,$$

and the other letters have the definitions previously given.

The Fourier cosine transforms calculated for each autocorrelation of the tide height record are

$$\begin{aligned} V_0 &= \frac{1}{(2)(12)}\left[1.32 + 2\sum_{q=1}^{m-1} C_q \cos\frac{q(0)\pi}{m} + 0.798 \cos(0)\pi \right] \\ &= \frac{1}{24}\Big[1.32 + 2(1)\ (-.312 - .553 + 1.20 - 0.154 - 0.735 \\ &\quad + 1.10 + 0.0476 - 0.917 + 0.921 + 0.252 - 0.998) \\ &\quad + 0.798(1) \Big] \\ &= 0.0765 \end{aligned}$$

$$\begin{aligned} V_1 = {} & \frac{1}{12}\left[1.32 + (0.798)\cos(1)\pi \right] + \frac{2}{12}\Big[-(0.312)\cos\frac{(1)(1)\pi}{12} \\ & -(0.553)\cos\frac{(2)(1)\pi}{12} + (1.20)\cos\frac{(3)(1)\pi}{12} - (0.154) \\ & \cos\frac{(4)(1)\pi}{12} - (0.735)\cos\frac{(5)(1)\pi}{12} + (1.10)\cos\frac{(6)(1)\pi}{12} \\ & + (0.0476)\cos\frac{(7)(1)\pi}{12} - (0.917)\cos\frac{(8)(1)\pi}{12} + (0.921) \\ & \cos\frac{(9)(1)\pi}{12} + (0.252)\cos\frac{(10)(1)\pi}{12} - (0.998)\cos\frac{(11)(1)\pi}{12} \Big] \end{aligned}$$

$$V_1 = 0.101$$

Similarly, the Fourier cosine transforms for the remaining autocorrelations can be computed:

$V_2 = -0.0332$
$V_3 = 0.0570$
$V_4 = -0.0443$

$V_8 = 0.995$
$V_9 = -0.224$
$V_{10} = 0.146$

$V_5 = 0.0943$ $V_{11} = -0.118$
$V_6 = -0.136$ $V_{12} = 0.0561$
$V_7 = 0.353$

Step 4. The final step in the spectral analysis of a single record is another weighting operation that counteracts some distortion of the spectrum resulting from the small sample size.

This step can be expressed mathematically as

$$U_0 = 0.54\,[V_0+V_1], \tag{3}$$

$$U_r = 0.23V_{r-1}+0.54V_r+0.23V_{r+1}, \tag{4}$$

$$\text{for } r = 1,2,3,\ldots m-1$$

$$U_m = 0.54\,V_{m-1}+0.54\,V_m, \tag{5}$$

where U_0, U_r, U_m are the power spectrum estimates corresponding to the

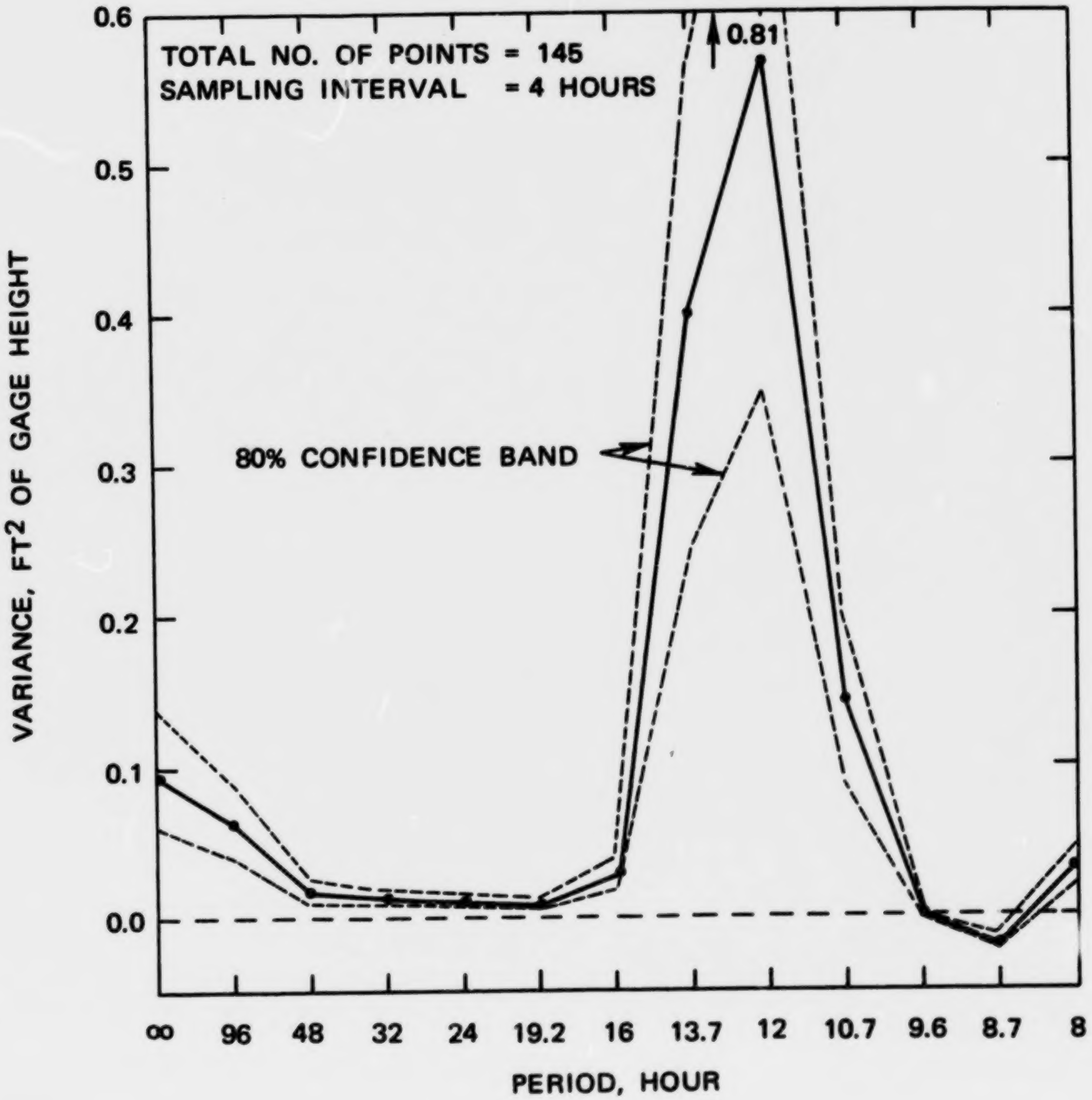


Figure 4. Spectrum of a Tidal Height Record.

respective lags, and the remaining symbols have the meanings previously assigned.

The power spectrum estimates for the tide gage record are

$$U_0 = (0.54)(0.0765)+(0.54)(0.101) = 0.0959$$

$U_1 = 0.0643$
$U_2 = 0.0183$
$U_3 = 0.0130$
$U_4 = 0.0109$
$U_5 = 0.00947$
$U_6 = 0.0295$
$U_7 = 0.388$
$U_8 = 0.567$
$U_9 = 0.142$
$U_{10} = 0.000446$
$U_{11} = -0.0173$
$U_{12} = -0.0334$

Each of these spectral estimates represents the part of the total record variance that is estimated to occur with a certain periodicity. The period corresponding to each lag is determined from the lag number and the sampling interval by this relation:

$$T_r = \frac{2m\ \Delta\tau}{r}, \tag{6}$$

where

T_r = period corresponding to lag r,
$\Delta\tau$ = sampling interval.

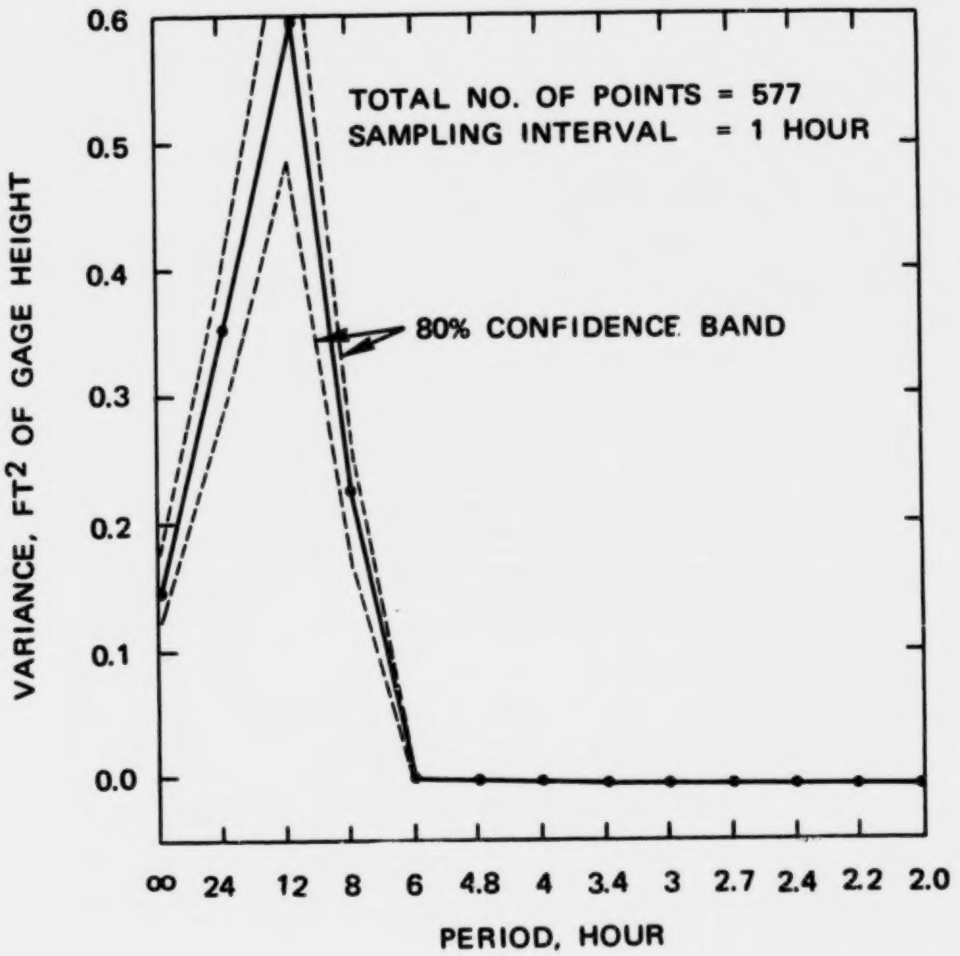
For the record analyzed here, $m = 12$ and $\Delta\tau = 4$ hours. The periods corresponding to the lag numbers are these:

Lag Number	*Period (hours)*
0	∞
1	96
2	48
3	32
4	24
5	19.2
6	16
7	13.7
8	12
9	10.7
10	9.6
11	8.7
12	8

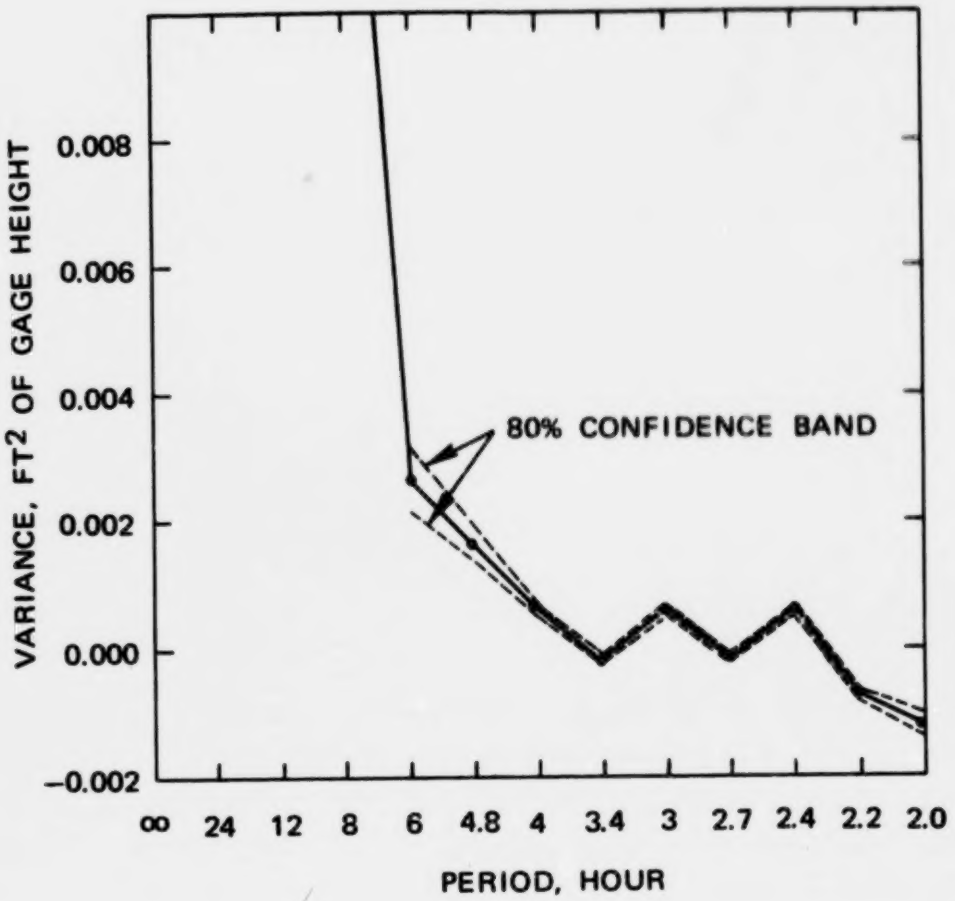
The spectral estimates are plotted as functions of period in Figure 4.

From these results is can be seen that about 79 percent of the observed total variance in water level can be attributed to periodicities of around 12 hours, that there are very small diurnal effects in the tide at this point, and that long-period changes account for about 12 percent of the observed variance.

The effects of using different sampling intervals and computing to different numbers of lags may be examined by considering the results of further spectral computation with this record. Figures 4, 5, and 6 illus-

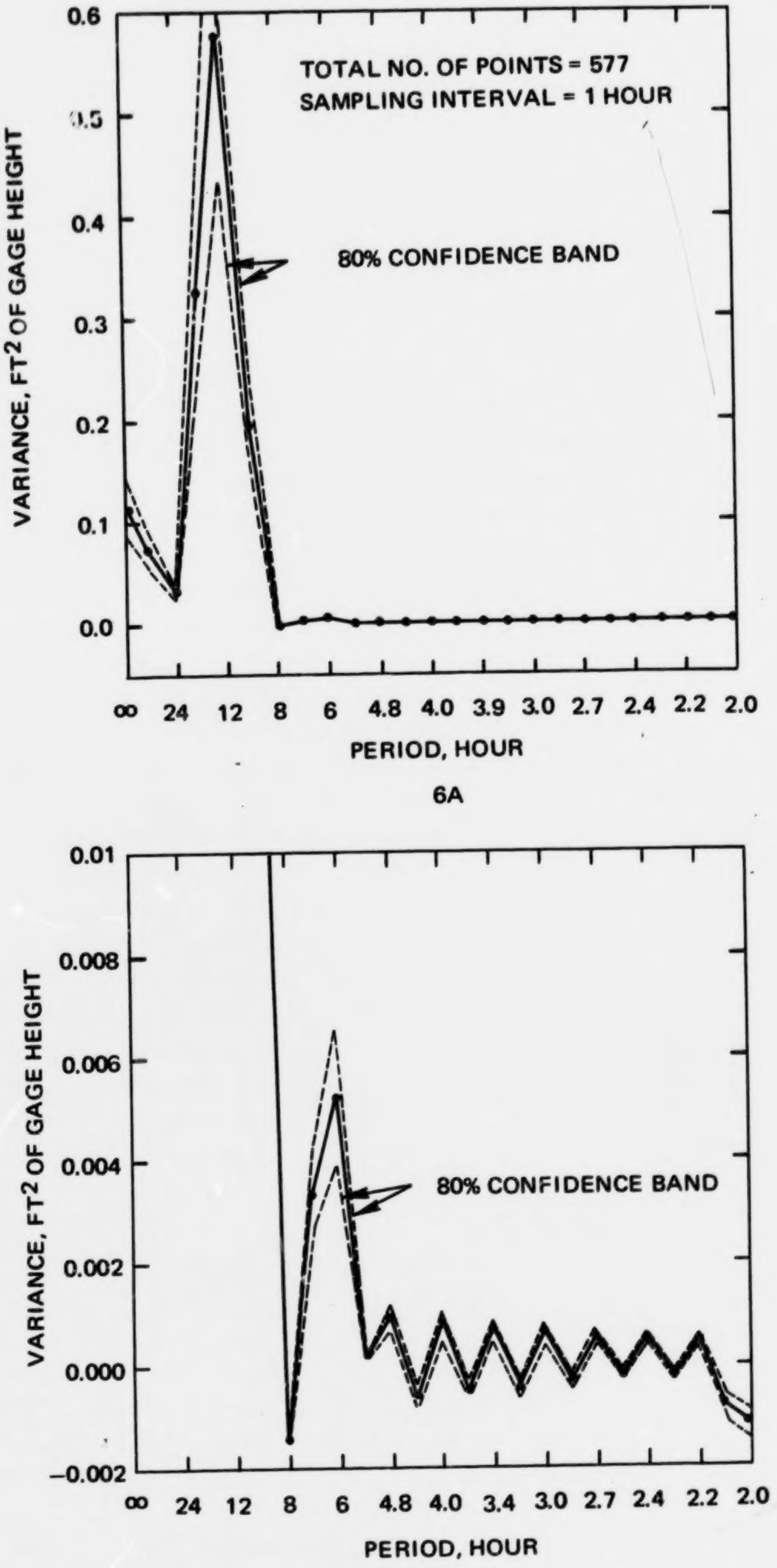


5A



5B

Figure 5. Spectrum of a Tidal Height Record.



6B

Figure 6. Spectrum of a Tidal Height Record.

trate spectra obtained from the same record, with variations only in the sampling interval and in the number of lags used in calculation. The total record length in each case was 576 hours. The spectral estimates from which Figure 4 was plotted were obtained by computation to 12 lags from values read at 4-hour intervals (145 points). Figure 5 was obtained from computation to 12 lags from values read at hourly intervals (577 points). Figure 6 was obtained from computation to 24 lags from values read at hourly intervals. The abscissal scales in Figures 5b and 6b are 50 times greater than those in Figures 5a and 6a; this magnification shows short-period effects more clearly.

While each of these figures shows the dominance of approximately semidiurnal periodicities in this record, the effects of the different sampling intervals and the different numbers of lags can be seen also.

Comparison of Figures 4 and 5 shows the effects of a change in sampling interval on the spectrum of the tide height record. First, the use of a smaller sampling interval for a given record length increases the number of measurements used in the analysis and therefore increases the number of degrees of freedom upon which each estimate is based. This results in the smaller confidence band shown in Figure 5. Second, the change of sampling interval from 4 hours to 1 hour permits the resolution of components with periods as short as 2 hours in the latter case instead of the 8 hours possible in the former. However, the use of the same number of lags with the shorter sampling interval does not permit as high a degree of resolution of long-period phenomena as was obtained with the longer sampling interval. In fact, the spectrum calculation leading to Figure 5 does not permit an estimate of the diurnal and longer-period effects because these are effectively masked by the dominant 12-hour component.

Comparison of Figure 5 with Figure 6 shows the effect of increasing the number of lags without changing the sampling interval. The resolving power is increased, *i.e.,* estimates are made for a greater number of spectral bands, but with a slight loss of confidence in the estimates, as is indicated by the wider confidence band in Figure 6 than in Figure 5. In this particular case, however, the increased resolving power demonstrates the existence of a small but statistically significant overtide (a harmonic of the semidiurnal tidal component) with a period of about 6 hours. This overtide is not shown in Figure 5b because the dominating semidiurnal band has overlapped the 6-hour-period band sufficiently to mask the very small overtide. In Figure 6b the resolving power of the 24 lags used in computation produces band-widths sufficiently narrow to prevent overlap of the 12- and 6-hour periods. The existence of higher harmonics is also shown in both Figures 5b and 6b. The former figure shows significant fourth and fifth harmonic overtides, whereas the latter shows these and several other short-period components.

The results of the spectral analysis of this record may be used to point out several characteristics of this technique.

First, each of the spectral values obtained represents an estimate of the variance over a range of periods in the vicinity of the nominal value. The range of periods for which each estimate is computed is determined by the sampling interval and the number of lags used in computation. This range is called the "equivalent width" (W_e) of the spectral band and can be calculated in terms of frequency by

$$W_e = \frac{1}{m\ \Delta\tau}. \tag{7}$$

The spectral estimate reported is an average value for all periods in the band over the range W_e. For example, for the results presented in Figure 4, the spectral estimate reported as corresponding to a 12-hour period is actually an average for the range of periods from 10.7 hours to 13.7 hours. The overlapping of spectral estimates is illustrated in Figure 5, and Figure 6 demonstrates how this overlapping can be reduced by increasing the number of lags, thereby reducing the equivalent width of the band for each spectral estimate.

Second, the precision of each estimate is a function of the total number of samples and the number of lags used in computation. A method for estimating the number of degrees of freedom for each estimate and for establishing the confidence intervals has been presented by Blackman and Tukey. In this method, the process being measured is regarded as Gaussian and the degrees of freedom and confidence intervals are based on a Chi-square distribution. The 80 percent confidence bands indicated on Figures 4, 5, and 6 have been determined by this method. It has been previously noted that a decrease in the total number of samples or an increase in the number of lags causes a widening of the confidence band, which represents a loss of precision in each spectral estimate.

Third, if the sampling interval is not small enough to permit resolution of the shortest periods that contribute significant variance to the record, the short-period variance in the record is not lost but is reported at harmonics of the true period. For example, if this record, as analyzed in Figure 4, contained significant variance with a period of 4 hours, this may be shown in the computations as part of the spectral estimate for the 8-hour period. This occurrence is known as "aliasing" or "folding." In Figure 4, the appearance of a significantly large estimate (in terms of absolute magnitude) at the shortest period computed in the analysis indicates that there is probably some aliasing of the short-period variance. The use of a smaller sampling interval, as shown in Figure 5, eliminated the aliasing in this spectrum.

Fourth, the appearance of an occasional negative spectral estimate is an artifact of the computation and results from the use of a finite record length for spectral computations. If the record were infinitely long, there would be no negative spectral estimates; but, when an estimate is close to zero in magnitude, it may appear with either a positive or negative sign. This result is interpreted only as a very small quantity of variance, and no physical significance is attributed to the negative sign.

PRECEDING PAGE BLANK NOT FILMED.

3

INTERPRETATION OF SPECTRA

To illustrate how the computed spectra of a set of observations can be used to gain an insight into the structure of a river or estuarine system, some data obtained in a field survey of the Potomac Estuary near Washington, D. C., are presented here as a basis for discussion. Since the purpose of this discussion is to present a conceptual picture of how spectral results may be interpreted, the quantitative results obtained from this particular study are not presented at this time.

In Figures 7 and 8 are presented time series of dissolved oxygen (DO) and 5-day biochemical oxygen demand (BOD) measurements obtained at six stations in this estuary. In Figures 9 and 10 are the spectra computed from these records. Each record contained 145 points obtained at 4-hour intervals; computation was carried out to 12 lags. Figure 11 is a schematic representation of the station locations.

Each spectrum shown in Figures 9 and 10 consists of three major effects: long-period, diurnal (24-hour), and semidiurnal (12-hour). The estimates adjacent to the long-period, 24-hour, and 12-hour periods are affected by the strength of the dominating period and the overlapping of spectral bands, as was discussed earlier in relation to the tide gage record. The significance of each of these effects may be considered separately.

The effects reported in the computation as "long-period" or of "infinite" period include all components whose periods are too long to be resolved in the computation. Since in these computations there is overlapping of the 96-hour and "infinite" periods, this means that all components with periods longer than 48 hours are regarded as long-period components. The use of the infinity symbol in the figures is merely a convenience in plotting and should be interpreted as meaning "long-period."

The components reported as long-period may have one or more of these physical interpretations:

1. The record may be affected by the existence of periodic components that are too long to be resolved by the length of record available. For example, a record of air temperature for a length of several months might show a long-period effect that, from a record of several years duration, would appear as an annual cycle.

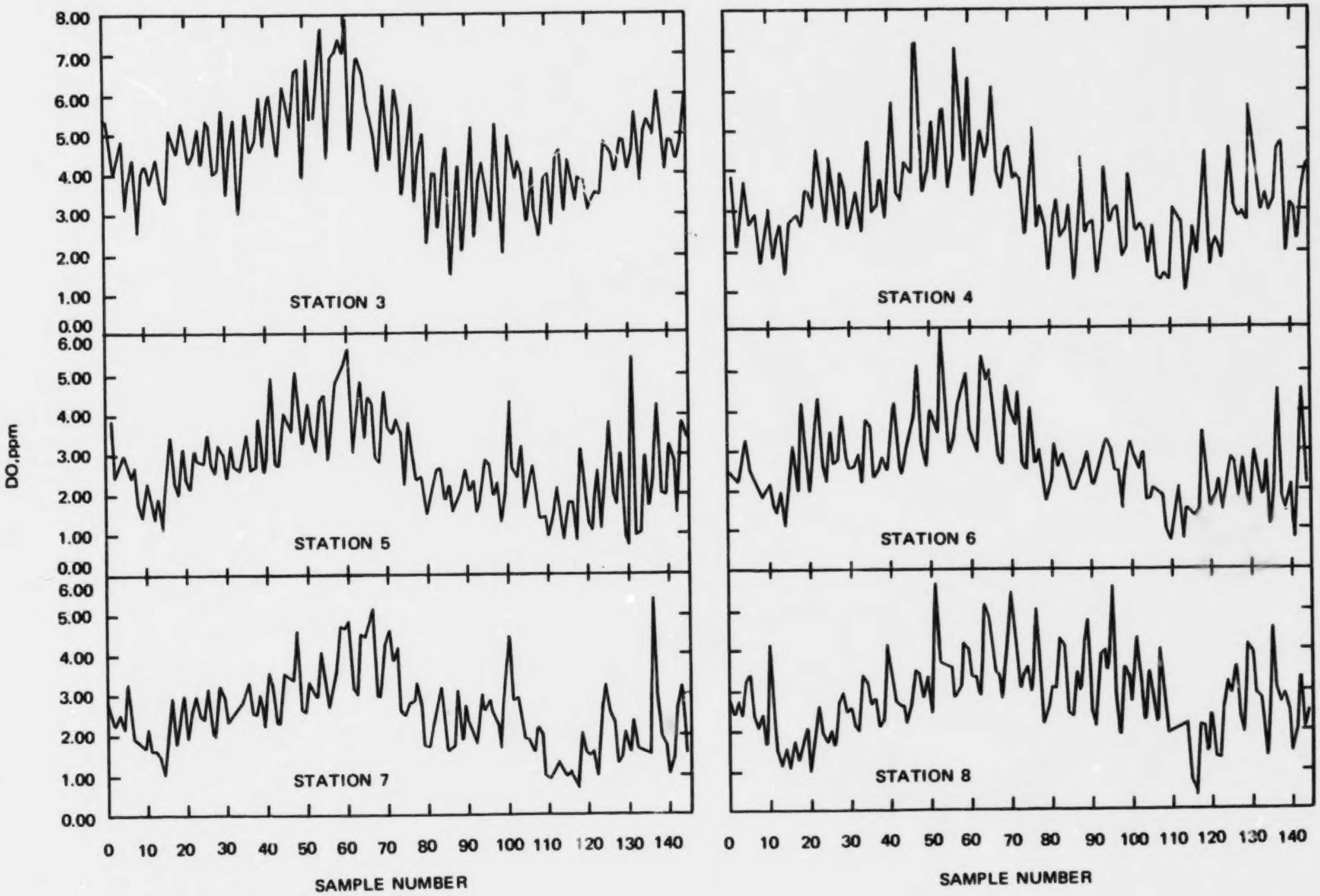


Figure 7. Dissolved Oxygen Records Obtained in the Potomac Estuary, August, 1959.

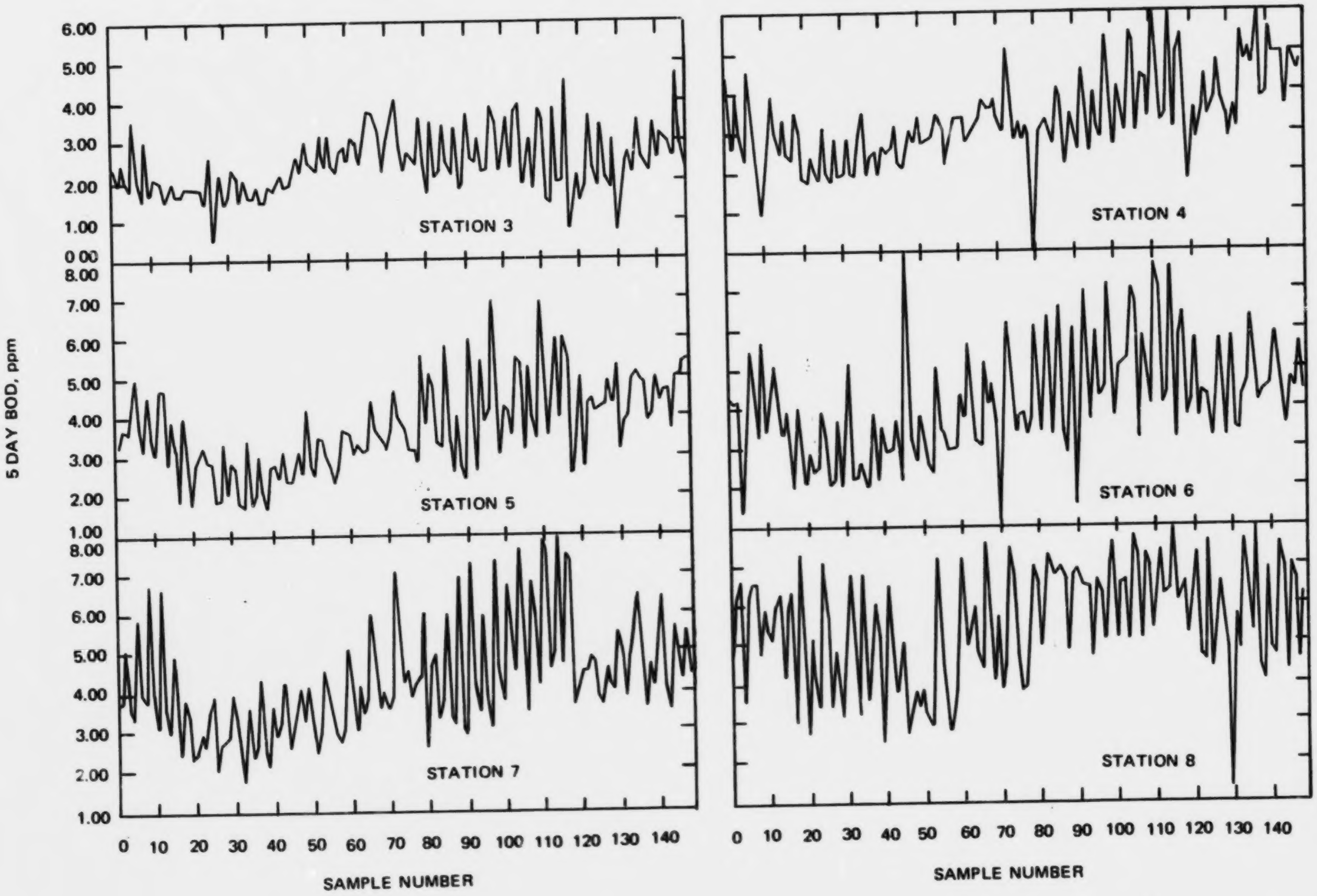


Figure 8. Biochemical Oxygen Demand Records Obtained in the Potomac Estuary, August, 1959.

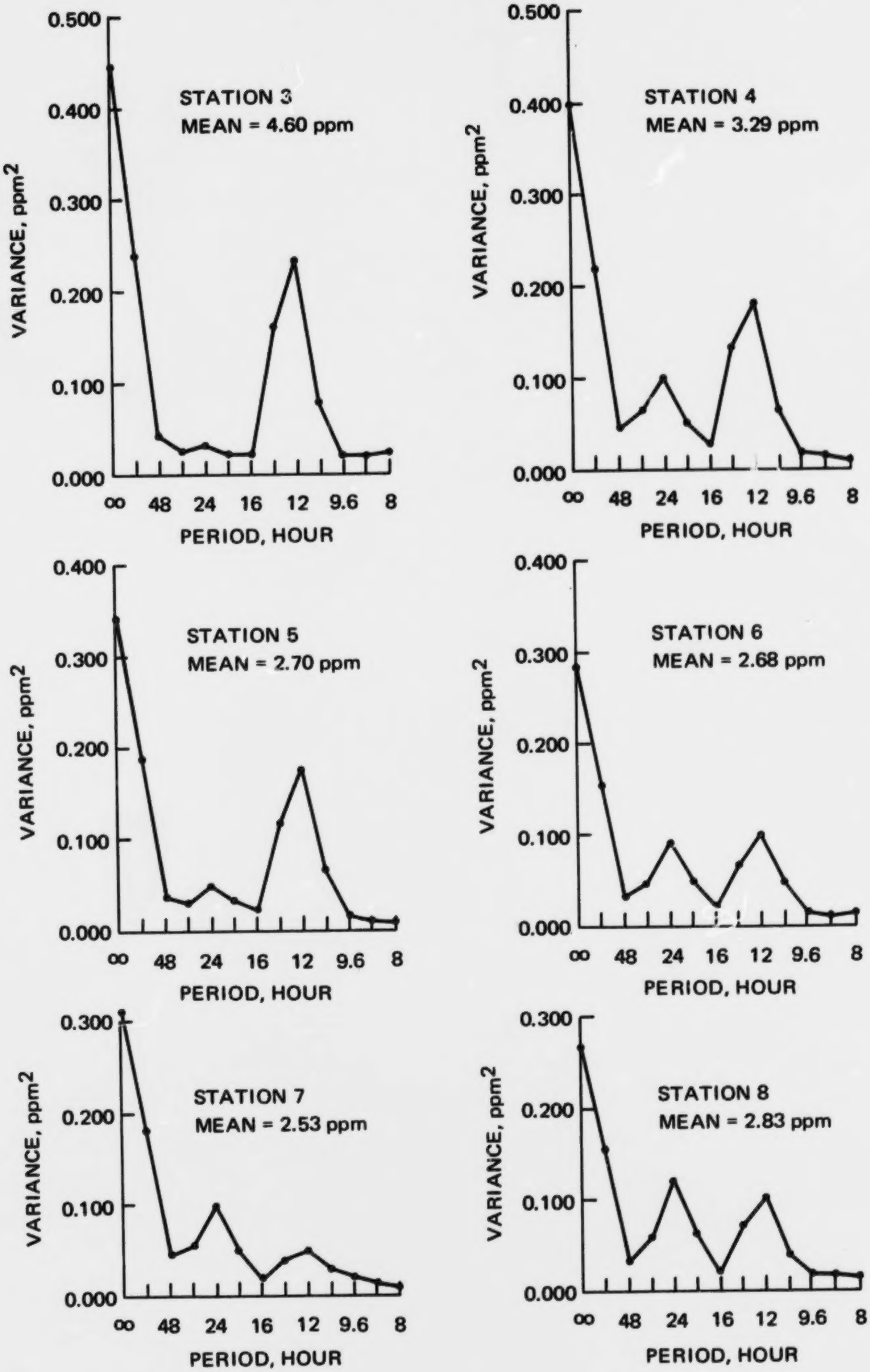


Figure 9. Dissolved Oxygen Spectra in the Potomac Estuary.

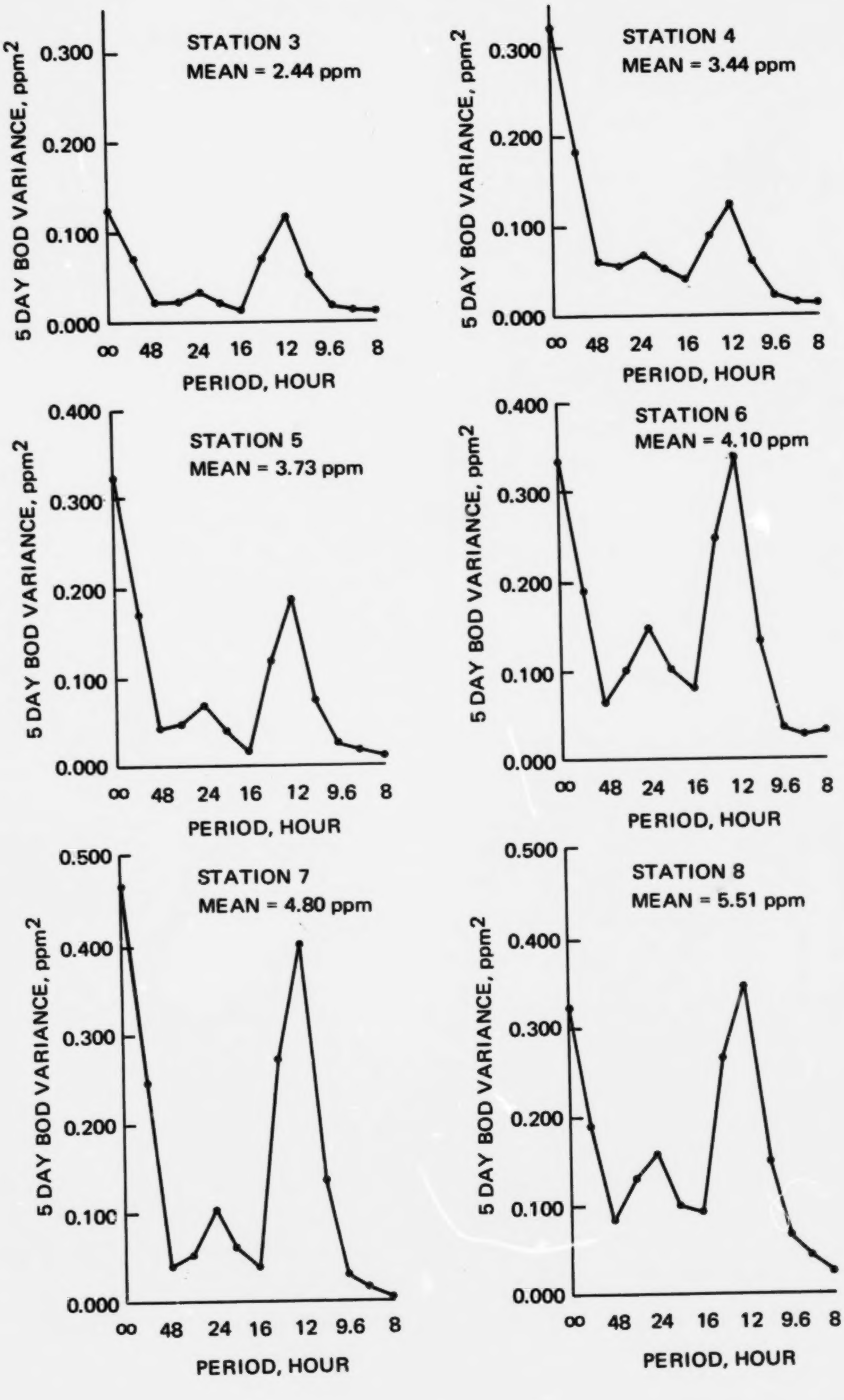


Figure 10. Biochemical Oxygen Demand Spectra in the Potomac Estuary.

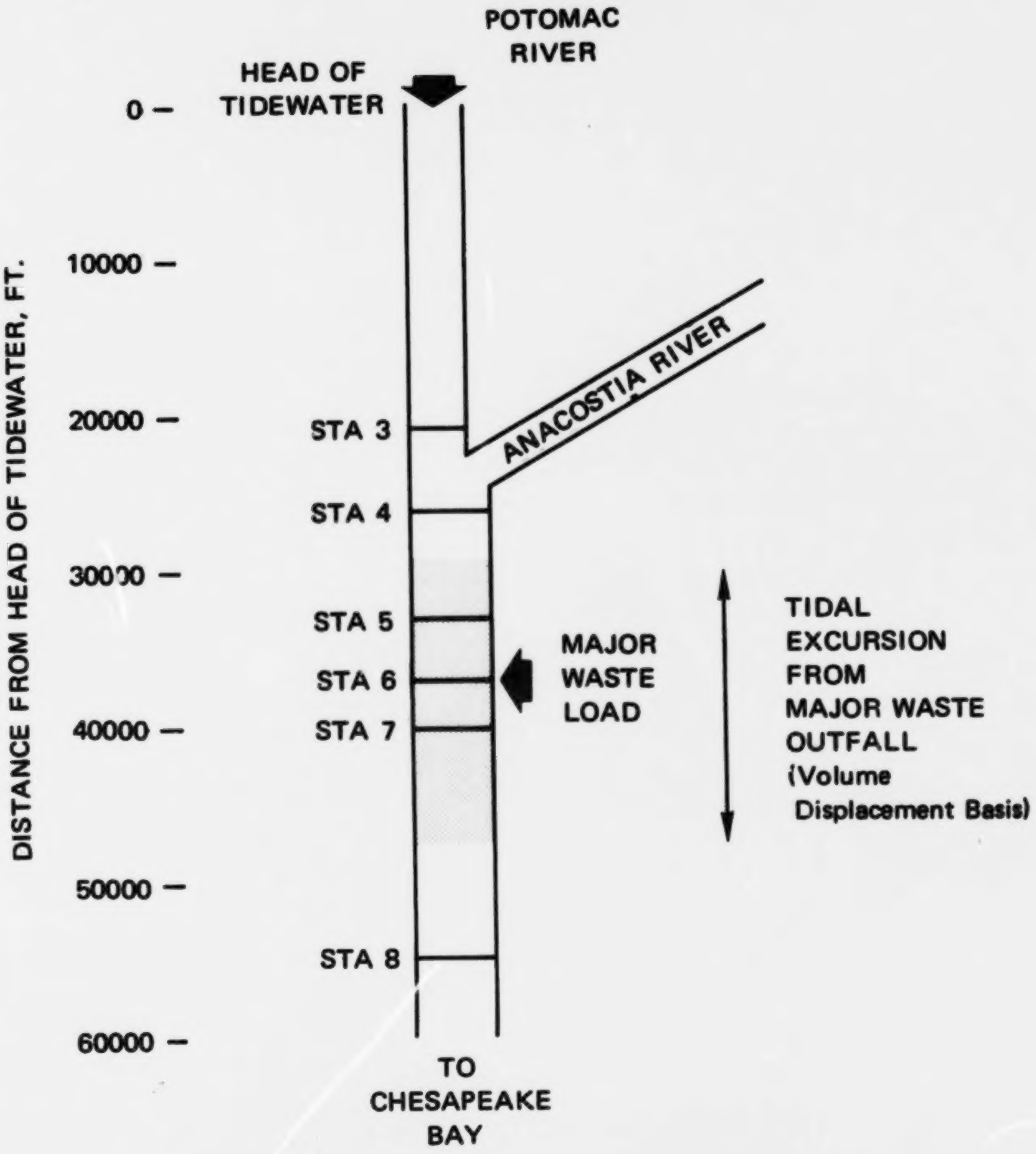

POTOMAC RIVER
HEAD OF TIDEWATER
0
10000
20000
30000
40000
50000
60000
DISTANCE FROM HEAD OF TIDEWATER, FT.
ANACOSTIA RIVER
STA 3
STA 4
STA 5
STA 6
STA 7
STA 8
MAJOR WASTE LOAD
TIDAL EXCURSION FROM MAJOR WASTE OUTFALL (Volume Displacement Basis)
TO CHESAPEAKE BAY

Figure 11. Schematic Representation of Potomac Estuary Sampling Station Locations.

2. There may be a secular trend in the record that is fundamentally aperiodic in nature. For example, a record of world population over several centuries might show such a trend.
3. Random sampling, reading, and analytical errors appear as long-period effects. A constant bias in the data affects the mean of the record only, and since spectra are computed from deviations from the mean, such a bias would not affect the spectral results.

From the computed spectra alone, it is not possible to differentiate among random errors, aperiodic effects, and long-period effects. In a practical sense it is often quite adequate to regard the long-period spectral estimates as secular trends for the available record length and to regard any random error as constant for all records of the same measurement.

The DO spectra presented in Figure 9 exhibit long-period effects that generally decrease from the upstream to the downstream stations. If a constant random error is assumed, the existence of an effect that is a function of distance from the head of tidewater may be postulated, and it might be deduced that this effect is related to the river discharge entering the estuary above Station 3. Since the parameter being measured is DO, this long-period effect might be interpreted as a measure of the amount of DO advected to the system in the river discharge or as an effect of river discharge on the reaeration characteristics of the system, or both. Considerable caution is in order if these results are to be interpreted in this manner; the change in the long-period estimate from Station 3 to Station 8 is barely significant at the 80 percent level, and the differences in this estimate between successive stations are not significant at the 80 percent level. The existence of a definite trend does indicate, however, that the interpretation given is reasonable and affords a basis for the examination of other results in terms of this hypothesis.

The BOD spectra in Figure 10 present an aspect somewhat different from the DO spectra at the same stations. The long-period effects at four of the six stations have very similar values.

Station 7 exhibits a long-period component that is somewhat higher than these but still lies within the 80 percent confidence band of the four stations. Station 3 exhibits a long-period effect that is significantly less than the corresponding effects at the other five stations. These components at the five downstream stations can be interpreted as the result of a secular trend in the major waste discharge to the system as well as random error in the determination of BOD.

Comparison of the long-period effects at Stations 3 and 4 with those of stations in the immediate vicinity of the waste outfall suggests that the limits of a tidal excursion upstream from the outfall may actually lie between Stations 3 and 4 instead of downstream from Station 4, as the calculations based on the tidal prism volume displacement indicate. The long-period effects at Station 3 might then be regarded as resulting from the longitudinal mixing of the waste discharge, whereas the similar effects at the other stations may be regarded as reflecting a combination of advective and diffusive processes.

The diurnal components of the DO and BOD spectra may be interpreted as expressing the effects of diurnal variations in waste discharge and in photosynthetic activity of the planktonic population. Diurnal variations in waste discharge would affect stations within a tidal excursion of the outfall more strongly than those beyond this distance, whereas photosynthetic activity in the system would be more pronounced at those stations exhibiting the higher nutrient concentrations, generally reflected in higher mean stream BOD's. Interpreted on this basis, the diurnal spectral components suggest that the extent of the upstream tidal excursion is between Stations 3 and 4, whereas Stations 6, 7, and 8 are subject to considerable photosynthetic activity in addition to the diurnal waste load variations. The diurnal components of the DO spectra at each station correspond in size to the respective BOD components; this is a result that might be expected from the theory of DO-BOD relationships in streams.

The semidiurnal component of these spectra reflects the advective motion of DO and stream waste load due to tidal action. The magnitude of this component at each station is a measure of the longitudinal concentration gradient of each parameter within a tidal excursion of the station. The DO spectra show large semidiurnal effects at Stations 3, 4, and 5, and relatively smaller ones at Stations 6, 7, and 8; these results indicate that there is a large DO gradient in the upper part of the system and a small one in the lower part. Examination of the mean DO values at each station suggests the existence of an oxygen-sag regime in which Stations 3 and 4 represent a zone of rapid degradation and Stations 5, 6, 7, and 8 a zone of critical DO and the beginning of recovery, a situation that is in close agreement with the spectral results. From this point of view the region of critical DO is the location that has the smallest semidiurnal spectral component, in this case Station 7.

The semidiurnal BOD spectral components present a picture somewhat different from the DO spectra. At the two upstream stations this component of the BOD spectrum is small. There is a significant increase in this component at Station 5 and again at Station 6, whereas the semidiurnal effects at Stations 7 and 8 are of the same magnitude as that at Station 6. These results suggest a low BOD gradient in the upper reaches of the system and a high gradient in the lower reaches, and Station 5 represents a region of transition. The strong gradients near and below the waste outfall suggest that the semidiurnal spectral components are affected by variations in the initial mixing of the waste load throughout the volume of water passing the outfall in a tidal excursion. It is apparent that the mean BOD's do not show such changes in gradient along the estuary. Comparison of the BOD gradient regime shown by the spectral analysis with that shown by the mean BOD values at each station illustrates the sensitivity of spectral analysis as a tool in estuarine engineering.

It has been the purpose of this discussion to demonstrate how spectral results can be interpreted in terms of familiar sanitary engineering concepts. It is not intended to suggest that the spectra can supply information in addition to that discussed. Spectral analysis is purely and simply a tool for the analysis of time-series data. It provides by itself no theoretical insight into natural processes, but it does permit one to examine individual periodic components of the data, with a minimum of interference from other components. As with any other statistical technique, the final interpretation of these results must be based on an understanding of the natural process, not on some magic numbers produced by the manipulation of data.

4

DESIGN CRITERIA FOR SPECTRAL ANALYSIS

In the design of any spectral analysis program the requirements for precision in each spectral estimate, for resolution of sufficient spectral bands, and for eliminating aliasing, or folding, must all be considered and balanced against each other before the sampling program and data analysis are begun.

An acceptable balance among these requirements can be established by careful choice of the sampling interval, the record length, and the number of lags used in computation.

The choice of appropriate sampling and computational factors should be based on reasonable assumptions of what the dominant periodicities in the system are. The shortest period it is necessary to resolve determines the sampling interval required, whereas the longest period necessary determines the total record length. In most sanitary engineering applications, it may be assumed that diurnal fluctuations in waste discharges and in photosynthetic activity will be of considerable importance. In tidal systems there will be significant semidiurnal periodicities; there may also be some effects of this short a period in waste discharge and in the biological regime for a nontidal system. As a general basis for experimental design, it may be assumed that resolution of 24-hour periods will be required in a nontidal system and that resolution of 12-hour periods will be required in a tidal system.

The shortest period it is theoretically possible to resolve with a given sampling interval is the period that is twice the sampling interval. That is, with a sampling interval of 6 hours, it is theoretically possible to resolve a 12-hour period. From a practical standpoint it is not possible to do this, since the 12-hour estimate would be the shortest period computed, and there would arise the question whether this is a valid estimate of a 12-hour period or whether it merely represents the aliasing of periods shorter than 12 hours. It is advisable that the sampling interval chosen be small enough to provide spectral estimates for several periods shorter than the expected dominant shortest period, so that this period is minimally affected by any aliasing that might occur.

As a rule of thumb, it is suggested that a maximum sampling interval of 8 hours is required to resolve 24-hour periods and an interval of 4 hours to resolve 12-hour periods. In general, a sampling interval of not more than one-third the length of the shortest significant period is recommended.

There are no clear criteria for determining the longest period that can be resolved from a given record. The longest period resolved in any particular analysis (other than the "infinite" period estimate) is determined by the number of lags used in computation and by the sampling interval. In general, computation to a number of lags greater than 10 percent of the total number of measurements in the record is not recommended, *i.e.,* for a record of 140 measurements, computation to no more than 14 lags is recommended. Each additional lag used in computation reduces the precision of all spectral estimates computed, and it is generally regarded that the 10-percent value affords an optimum balance between precision of individual estimates and resolution of spectral components. As a guide in determining the required record length for design purposes, it may be assumed that a record length at least 10 times as long as the longest significant period to be resolved will be required. For many sanitary engineering field surveys, the 24-hour period is about the longest significant period it is necessary to resolve from field survey data; in such cases a minimum record length of 240 hours would be required.

As a basis for selecting the number of lags to be used in computation 10 percent of the number of measurements in the record is used as an upper limit. The number of lags finally chosen will probably be based on the resolution believed to be required. For example, if it is desired to separate diurnal and semidiurnal components, the number of lags chosen must be large enough to include several estimates between the 24- and 12-hour estimates, so that there is a negligible amount of interference between the major components. An acceptable degree of precision is then obtained by increasing the record length if the numbers of degrees of freedom on which each estimate is based give a confidence band that is too large to give usable results.

Any type of time-series record can be subjected to spectral analysis if it represents sampling at uniform time intervals and if there are no missing points; if a few points are missing, however, a limited amount of interpolation may be done. Interpolation of the mean value of all the measurements or linear interpolation of a missing point between two measurements is the usual approach. There are no definite criteria that serve as an indication of how much interpolation can be done in any particular case; a general consideration of the process suggests, however, that if the missing points are widely scattered, up to possibly 5 percent of the data may be interpolated without serious effects on the computed spectra.

Data suitable for spectral analysis are obtained most simply and cheaply by means of automatic sampling and recording equipment. Conversely, spectral analysis offers the most effective means of analyzing and correlating the large quantities of data produced by such instruments. This does not mean, of course, that the computation of spectra is limited to time-series data obtained from automatic devices. The DO and BOD data discussed here were obtained by conventional manual sampling procedures, whereas the tide height record was obtained from an automatic recording device. If the requirements of uniform sampling interval and record length are met, the means by which the data are obtained is immaterial.

The computation of spectra is most cheaply and efficiently accomplished by high-speed digital computer. The spectra presented here were computed on an IBM 704; the total cost, including preparation of the data for the computer, was estimated at less than 10 dollars per spectrum. Computation by hand for all but very short record lengths is prohibitive from the standpoint of time, money, and accuracy.

PRECEDING PAGE BLANK NOT FILMED.

5

INTRODUCTION TO CROSS-SPECTRA

The discussions in the preceding chapters were concerned entirely with the generalized harmonic analysis of a single data record. The end product of such an analysis is a spectrum of magnitudes of variance, showing the frequencies and magnitudes of the dominant periodic components of the record.

Now consider the spectral analysis of a pair of records analyzed together. The computations are quite similar, but the use of two records instead of one permits the calculation of some additional statistics, which have some quite direct and highly significant engineering uses. The means by which these statistics may be calculated it shown in some detail later; at this time let us briefly outline the types of information expected from cross-spectrum calculations and indicate what these may mean as aids in determining engineering parameters of estuarine systems.

In the spectral analysis of a single record all phase relationships are lost. For example, a record of dissolved oxygen (DO) might be affected by several different effects, which vary over a diurnal period. The diurnal change in water temperature might have an effect on DO, and the change in photosynthetic activity of algae may also have an effect on DO. Both of these effects would be diurnal in frequency, but the peak of DO due to photosynthesis could occur at a different time from that reflecting the change in water temperature. With but a single record to work from, only the total magnitude of the resulting diurnal variation in DO can be determined. The phase relationships between these two diurnal effects cannot be established from a single record.

One of the powerful consequences of the cross-spectrum calculation is the maintenance of phase relationships between dominant periodic components of the two records being analyzed. Cross-spectral analysis of simultaneous water temperature and DO records would show by how many hours the daily peak in DO follows the peak in daily temperature and the size of the variation. (Since these two variables are inversely related, the phase lag would show that they are out of phase, *i.e.,* the phase lag would be greater than 180 degrees.) Similarly, cross-spectral analysis of simultaneous sunlight and DO oxygen records would permit the calculation of phase lag relationships between sunlight intensity and photosynthetic oxygen production.

The use of two records simultaneously results in the computation of four sets of spectra: the individual power spectrum of each of the two records, and two spectra resulting from the two cross-correlations of the two records. These two spectra, designated the *cospectrum* and *quadrature spectrum,* are components of the cross-spectrum itself. (The relationship of these two components to each other and to the cross-spectrum will be discussed in detail.)

For each of the frequency bands for which spectral density estimates are made, there therefore result four estimates of spectral density reflecting the relationships of two parameters. From these four estimates several other extremely useful statistics can be calculated.

Coherence is analogous to the square of a correlation coefficient.

Phase lag is the angular time lapse between respective maxima or minima in the two records.

Response function is a quantitative estimate of the amount of variation in one record that is associated with similar variation in the other record.

These statistics are extremely easy to interpret in terms of the physical environment through standard forms of mathematical analysis.

Coherence has many of the properties of the square of a correlation coefficient and is often interpreted as such. In magnitude it is expressed on a 0 to 1.0 scale and is dimensionless. The coherence measures the closeness with which changes in one record follow those in the other. This closeness may be in both magnitude of variation and the closeness with which the frequencies correspond. If the one record is affected by several driving forces with the same dominant periodicities, the coherence may reflect that fraction of the variation in one record which is associated with the other. In concept, then, coherence tells us how closely the two records are related.

Phase lag is a measure of the time needed for changes in one record to bring about corresponding changes in the other. It is computed in terms of radians but can easily be converted to hours, days, or any other convenient time unit. So, to go with the coherence, there is a measure of time relationships between two parameters.

The response function is the spectrum of one record computed from the spectrum of the other record and from the cospectrum and quadrature spectrum of the two records. It is expressed in units of the original record and shows the total magnitude of the variance in the one record that is associated with the other.

These three statistics, then, define three important factors in relating two parameters to each other. First, a measure of how closely associated they are: the coherence; second, a measure of their dynamic (or timewise) relationship to each other: the phase lag; and third, a measure of the magnitude of the association: the response function.

Examination of the relationship of one set of data to another is almost always concerned with establishing some sort of cause-and-effect relation. This is true in cross-spectral analysis as well as other types of analysis. Accordingly, one of the records to be subjected to cross-spectral analysis is termed the *input* and the other is called the *output*.

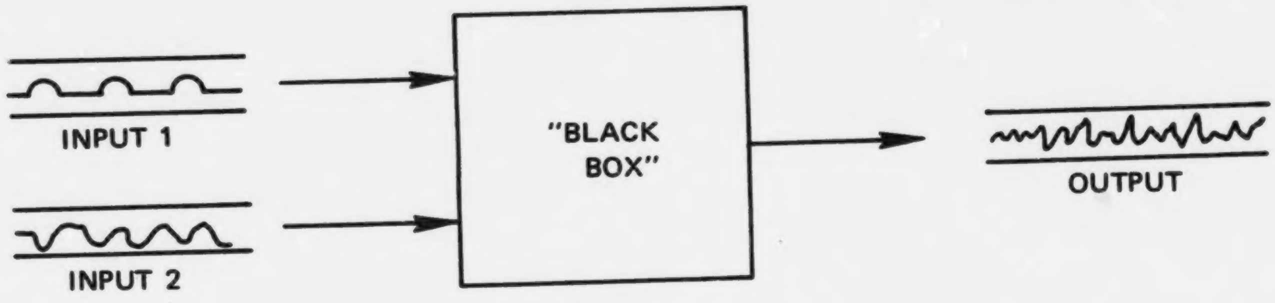


Figure 12. Typical Input-Output System

The use of these terms is to regard the environment as a "black box." That is, it is not possible to state precisely what is happening in the environment; but it is possible to measure what goes in and what comes out, and from these measures deduce a little about what is happening in the enigmatic "black box" (Figure 12). In computing cross-spectra, then, it is assumed measurements are made of one causative factor and one resulting condition, and that one record is then an input and the other one an output.

This terminology is a convenience in interpreting the results and a proper choice of input and output records can usually be made on physical grounds. For example, in an analysis of a temperature record and a DO record, the temperature is the input and the DO is the output.

The mathematics themselves are not affected by this choice, but the proper choice of input and output will avoid the trauma of trying to visualize negative phase lags, for example.

PRECEDING PAGE BLANK NOT FILMED.

6

COMPUTATION OF CROSS-SPECTRA

The general scheme for the calculations of cross-spectra is the same as that for the computation of individual power spectra. There are some differences in computation and additional complexity of notation brought about by the use of two records instead of one.

From an operational standpoint, the calculation of cross-spectra is much more involved than the computation of individual power spectra because the cross-spectrum scheme of analysis not only includes the computation of four spectra in each analysis but also includes further operations with the resulting spectra. For this reason no attempt is made to follow through an actual arithmetic example to its conclusion. Instead a very simple algebraic example is used to illustrate the differences in computation and approach which are required in this scheme of analysis.

Suppose that there are two records of data to subject to spectral analysis. In order to avoid having to carry a mean value throughout the discussion assume that the arithmetic mean is zero for each record. Let

A = a record of 100 temperature observations at uniform time intervals, and

B = a record of 100 dissolved oxygen observations of the same system over the same uniform time intervals.

The individual sequential values of each record may be identified as $a_1, a_2, a_3, \ldots a_{100}$, and $b_1, b_2, b_3, \ldots b_{100}$.

As the initial step in the analysis, form the autocorrelation of each record and the two cross-correlations of the pair of records. The autocorrelations for each record at lags 0 and 1 will follow this form:

$$C_0(A) = (a_1 \cdot a_1 + a_2 \cdot a_2 + a_3 \cdot a_3 + \ldots + a_{100} \cdot a_{100}) \frac{1}{100}, \tag{8}$$

$$C_1(A) = (a_1 \cdot a_2 + a_2 \cdot a_3 + a_3 \cdot a_4 + \ldots + a_{99} \cdot a_{100}) \frac{1}{99}, \tag{9}$$

$$C_0(B) = (b_1 \cdot b_1 + b_2 \cdot b_2 + b_3 \cdot b_3 + \ldots + b_{100} \cdot b_{100}) \frac{1}{100}, \tag{10}$$

$$\mathrm{C}_1(B) = (b_1 \cdot b_2 + b_2 \cdot b_3 + b_3 \cdot b_4 + \ldots + b_{99} \cdot b_{100}) \frac{1}{99}. \tag{11}$$

From the pair of records the cross-correlations are formed. Here enters a complexity. Look briefly at how the cross-correlations are calculated. At lag 0, it is quite straightforward; the sequential values of each record are multiplied together, so that there is something of the form

$$a_1 \bullet b_1 + a_2 \bullet b_2 + a_3 \bullet b_3 + \ . \ . \ . \ + a_{100} \bullet b_{100}$$

When the first lagged product is computed, however, there are two possible sequences of products which may result:

(1) $a_1 \bullet b_2 + a_2 \bullet b_3 + a_3 \bullet b_4 + \ . \ . \ . \ + a_{99} \bullet b_{100}$

(2) $b_1 \bullet a_2 + b_2 \bullet a_3 + b_3 \bullet a_4 + \ . \ . \ . \ + b_{99} \bullet a_{100}$

It is immediately apparent that the sums of these two sequences of products will be the same only when the two records have identical values. When there are two different records, then, there are two cross-correlations. The cross-correlation functions used in the cross-spectrum scheme of data analysis actually consist of the sums and differences of the two basic cross-correlations.

The general equations for these two cross-correlation functions are

$$S_r{+} = \frac{1}{2(n-r)} \sum_{t=1}^{n-r} (x_t y_{t+r} + x_{t+r} y_t), \tag{12}$$

$$S_r{-} = \frac{1}{2(n-r)} \sum_{t=1}^{n-r} (x_t y_{t+r} - x_{t+r} y_t), \tag{13}$$

where x and y denote values of the records being used,

t = 1,2,3, . . . ,n = sequential index of values,
r = 0,1,2, . . . ,m = lag number,
m = total number of lags.

For the example, at lag 0 there are the equations

$$S_0{+} = \frac{1}{2(100)} [(a_1 \bullet b_1 + a_1 \bullet b_1) + (a_2 \bullet b_2 + a_2 \bullet b_2) + \ . \ . \ . \\ + (a_{100} \bullet b_{100} + a_{100} \bullet b_{100})]$$

$$S_0{-} = \frac{1}{2(100)} [(a_1 \bullet b_1 - a_1 \bullet b_1) + (a_2 \bullet b_2 - a_2 \bullet b_2) + \ . \ . \ . \\ + (a_{100} \bullet b_{100} - a_{100} \bullet b_{100})]$$

For lag 1 there are the equations

$$S_1{+} = \frac{1}{2(99)} [(a_1 \bullet b_2 + a_2 \bullet b_1) + (a_2 \bullet b_3 + a_3 b_2) + \ . \ . \ . \\ + (a_{99} \bullet b_{100} + a_{100} \bullet b_{99})]$$

$$S_1{-} = \frac{1}{2(99)} [(a_1 \bullet b_2 - a_2 \bullet b_1) + a_3 \bullet b_2) - a_3 b_2) + \ . \ . \ . \\ + (a_{99} \bullet b_{100} - a_{100} \bullet b_{99})]$$

The next step in the analysis is to compute the Fourier transforms for the two autocorrelations and for the two cross-correlation functions. For the cross-correlation function $S_r{+}$, the Fourier cosine transform (C_r') is computed using the relationships

$$C_0' = \frac{1}{m}\left[\frac{1}{2}\,(S_0^+ + S_m^+) + \sum_{1}^{m-1} S_r^+\right], \tag{14}$$

$$C_r' = \frac{2}{m}\left\{\frac{1}{2}\,[S_0^+ + (-1)^r S_m^+] + \sum_{q=1}^{m-1} S_q^+ \cos\frac{\pi rq}{m}\right\}, \tag{15}$$

$$C_m' = \frac{1}{m}\left\{\frac{1}{2}\,[S_0^+ + (-1)^m S_m^+] + \sum_{r}^{m-1} (-1)^r S_r^+\right\}. \tag{16}$$

For the cross-correlation function S_r^-, the Fourier sine transform Q_r' is computed according to the relationships

$$Q_0' = 0, \tag{17}$$

$$Q_r' = \frac{2}{m}\sum_{q=\mathrm{v}}^{m-1} S_q^- \sin\frac{\pi rq}{m}, \tag{18}$$

$$Q_m' = 0. \tag{19}$$

This step performs the same smoothing function for the cross-correlation functions as the Fourier transformation does for the autocorrelation. The use of the cosine transform on one of the pair and the sine transform on the other is a key step in the calculation which permits the computation of phase lags from the results of the analysis.

The reason for this lies in the fundamental geometrical relationship of the sine and cosine. That is, in a given right triangle, the sine of one of the angles is the ratio of the opposite side to the hypotenuse, while the cosine is the ratio of the adjacent side to the hypotenuse. Numerical values of the sine and cosine are equal for angles which are 90 degrees out-of-phase with each other, so that the effect of multiplying one cross-correlation by the sine and the other by the cosine is to weight each cross-correlation by a function which suppresses components of the variance which are not 90 degrees out-of-phase with each other.

The final step in the treatment of the cross-correlation functions is a weighting process exactly analogous to that used in the computation of individual power spectra. The results consist of the cospectrum (C_r) and the quadrature spectrum (Q_r), which are computed using these relationships:

$$C_0 = 0.54\,(C_0' + C_1'), \tag{20}$$

$$C_r = 0.23\,C_{r-1}' + 0.54\,C_r' + 0.23\,C_{r+1}', \tag{21}$$

$$C_m = 0.54\,(C_{m-1}' + C_m'), \tag{22}$$

$$Q_0 = 0.54\,(Q_0' + Q_1'), \tag{23}$$

$$Q_r = 0.23\,Q_{r-1}' + 0.54\,Q_r' + 0.23\,Q_{r+1}', \tag{24}$$

$$Q_m = 0.54\,(Q_{m-1}' + Q_m'). \tag{25}$$

These two results consist of two spectra which are 90 degrees out of phase with each other. They may be considered components of the output spectrum. A crude geometrical analogy to what has been done is presented in Figure 13. Instead of a resolution of vectors into components, there is a resolution of spectra into components.

The four basic statistics computed in this analysis are these:

U_{ri} = power spectrum of input record
U_{ro} = power spectrum of output record
C_r = cospectrum
Q_r = quadrature spectrum

A number of other statistics can be calculated from these basic results. There are considered here three of these statistics which appear to have a basic relationship to studies of the dynamics and water quality of estuarine systems.

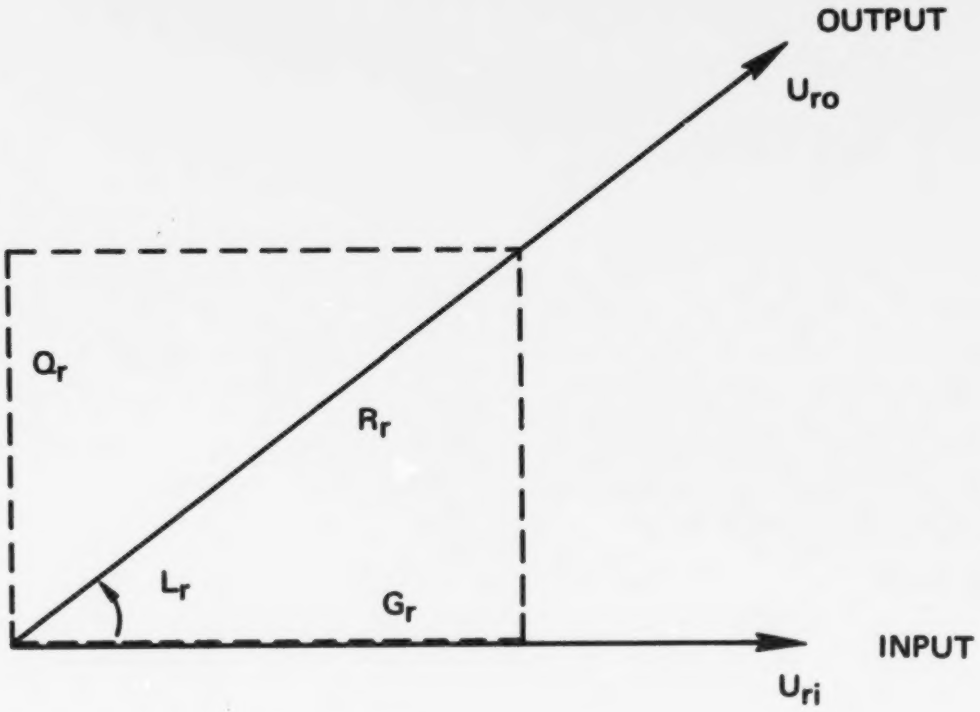


Figure 13. Geometrical Analogy to Cross-spectral Analysis

The *coherence* (H_r) is computed according to the relation

$$H_r = \frac{C_r^2 + Q_r^2}{U_{ri} \cdot U_{ro}}. \tag{26}$$

H_r is a dimensionless number, and is generally regarded as analogous to the square of a linear correlation coefficient. The terms in this equation are directly related to terms in standard formulations for a linear correlation coefficient. One such form is

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y} = \text{linear correlation coefficient of } y \text{ or } x$$

S_{xy} = covariance of y and x values
S_x^2 = variance of x values
S_y^2 = variance of y values

H_r represents a component-by-component expression for such a statistic, while r_{xy} represents an averaged value for the entire records of x and y values.

The *phase lag* (L_r)is computed by the relation

$$L_r = \text{arc tan} \left[\frac{Q_r}{C_r} \right]. \tag{27}$$

The units are radians, which can be converted into any convenient time units related to the sampling interval. Since Q_r is a sine transform and C_r is a cosine transform, the reason for this relationship is obvious.

The *response function spectrum* (R_r) is computed by the relation

$$R_r = H_r \bullet U_{ro} \tag{28}$$

and has the units of the output spectrum.

These three statistics in addition to the four spectral values form the basis for a coherent scheme of data analysis to interpret sampling results in terms of the physical environment.

PRECEDING PAGE BLANK NOT FILMED.

7

INTERPRETATION OF CROSS-SPECTRA

The purpose in subjecting data to any form of mathematical manipulation is to establish relationships between parameters of the physical environment. Any statistic is useful only when it can be interpreted in terms of the environmental situation. It is necessary, therefore, to relate the statistics computed in cross-spectral analysis to properties of the environmental system.

The interpretation of individual power spectra has already been discussed in some detail and needs no further comment here.

The cospectrum and quadrature spectrum are the basis for several further calculations showing the degree of correlation between two records. The actual values of the cospectrum and quadrature spectrum are right-angle components of the response of the output spectrum to the input spectrum. In terms of the physical environment, these components have no direct significance; their value and utility lie in the fact that from them can be calculated coherence, phase lags, and response spectra.

It is worth noting that the relative magnitudes of the cospectrum and quadrature spectrum govern the phase lag, and their absolute magnitudes dominate the coherence.

Coherence. The coherence is roughly analogous to the square of a correlation coefficient and, in the simplest use of spectral analysis, it is often interpreted exactly as a linear regression correlation coefficient is interpreted. Because of the means whereby it is calculated and because it is derived from environmental rather than isolated and controlled measurement, it may be interpreted in a number of other ways peculiar to environmental studies and to cross-spectral analysis.

Conceptually, the term "coherence" is precisely descriptive of what the statistic means. Coherence is a measure of how coherent the two records are; that is, how similar the variances of the two records are at each spectral estimate and how well the fluctuations in the one record follow those in the other. The coherence, therefore, is not really merely a measure of linear correlation of two variables; it has implications concerning both the magnitudes of variances observed and the shape of the spectral components examined.

This point may be clarified with an example. Consider two records for cross-spectral analysis: a record of sunlight intensity and a record of dissolved oxygen determination. The actual data records for several clear days might look like Figure 14.

The solar intensity record reflects the fact that, when the sun sets, the parameter immediately reaches its minimum value, while the DO record reflects the fact that there is a sequence of events and processes between the sun's energy and its ultimate effect on DO.

These two records are quite dissimilar in shape, not necessarily because of a lack of correlation between the two, but because of the nature of the processes involved. The coherence calculation reflects only the actual numbers in the two records, so the coherence computed between these two records might be a bit lower than the physical situation actually warrants. As with any other statistic, coherence must be interpreted with a clear realization of how the number was derived.

As a general rule of thumb in interpreting coherencies, it can be stated that a coherence is always lower than, or equal to, the true degree of correlation between two records.

One of the most useful ways to interpret coherence is to regard it as the fraction of the variance in the output record that is related to similar variance in the input record. For example, if the coherence between the diurnal components of the solar intensity and DO records was 0.60, we could state that about 60 percent of the diurnal variance in DO is explained by the diurnal component of variance in solar intensity.

This interpretation of coherence follows directly from its means of computation. The coherence, as calculated, is the ratio of the sum of the cross-covariances to the product of the individual power spectra; that is, the coherence is actually a ratio of squared variances. Its interpretation as a fraction of variance is therefore quite logical.

Response function spectrum. This statistic is the calculated response spectrum of one record to another. The response function spectrum shows what the output spectrum would be like if the input record were the only parameter dominating it. The response function spectrum has the same physical meaning as the individual power spectra have, and may be interpreted in much the same manner.

The response function spectra of one output to several different input parameters can be computed separately and the results compared to determine what major driving forces are affecting any particular parameter. Response function spectra do reflect the phase relationships between the input and output records; in this respect they may therefore be interpreted more advantageously than the power spectra of the individual records.

Phase lags. The phase lag between two records estimates the time delay in the process or processes which the two records measure. The physical significance of this delay may be interpreted in several ways depending on the nature of the records and processes.

For records of different parameters at a single station the phase lag may quantitatively estimate the time required for a biological, chemical, or physical process to occur. For example, a phase lag between air tempera-

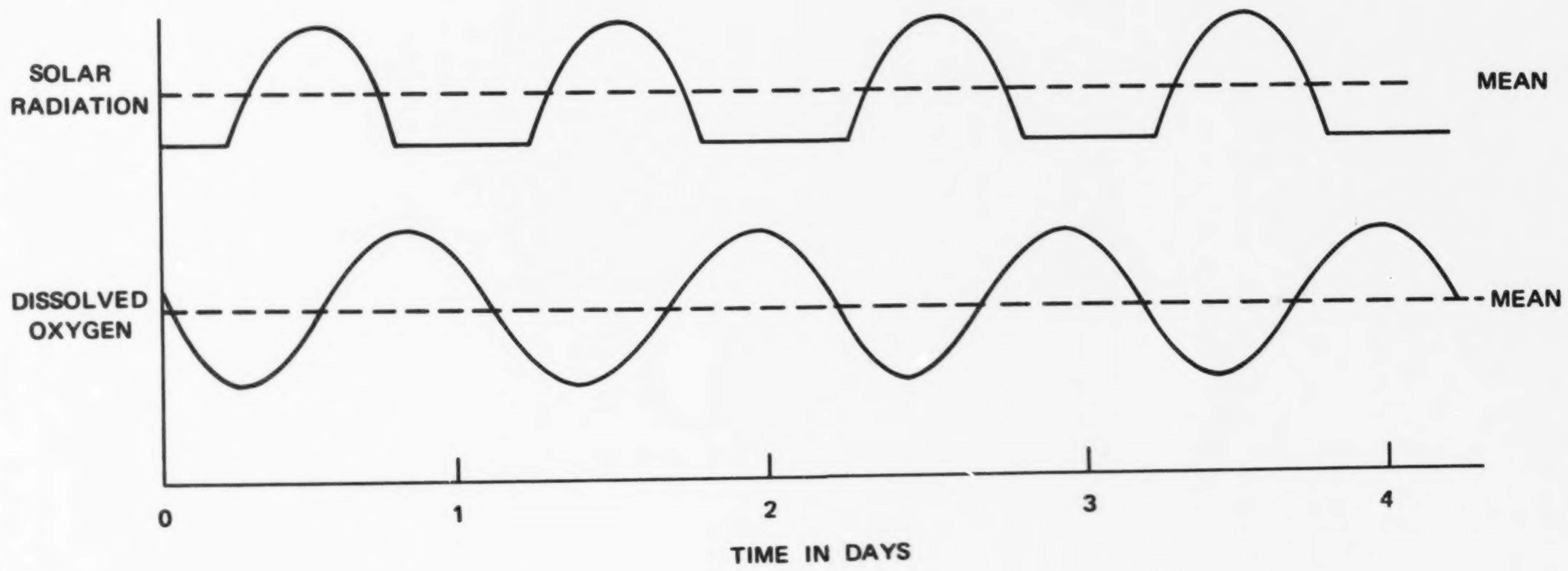


Figure 14. Comparison of Typical Data Records.

ture and water temperature would estimate the time required for the transfer of thermal energy from the air to the water or *vice versa*; the phase lag between solar intensity and DO would measure the time delay in the biological and chemical processes of photosynthetic oxygen production.

For records of the same parameter at different stations, the phase lag may be a measure of advective transport between two stations. For example, the phase lag of tidal height between two stations might show the progression of the tidal wave up an estuary. The phase lag between DO's at two stations might, however, measure the change in DO due to consumption or reaeration during the time of the phase lag as well as advective transport; a considerable degree of caution in interpreting phase lags in this manner is indicated when a combination of physical, chemical, and biological processes is possible.

In any form of interpretation of cross-spectral results the procedures of the calculation must be clearly understood and their significance kept firmly in mind.

The cospectrum and quadrature spectrum are variances calculated from the mean values of the records, just as the individual power spectra are calculated. This does not affect the values of coherence, phase lag, or response spectrum, but it can affect the interpretation that may be placed upon these values. For example, if the mean DO of a record is very near the value of DO saturation for the temperatures and salinities concerned, one would expect the autocovariance function and the cross-covariance functions to be smaller than they would be if the mean DO of the record was about 50 percent of saturation, because the DO saturation value is a limiting value for the record. The same is true if the mean value of the record is very near 0 percent of saturation.

The relationship of the mean value and limiting values of the recorded parameter to the record variance and other descriptive statistics may be illustrated with some data collected on the Potomac estuary.

In Figure 15 are shown mean values and variances of DO records with 145 sequential values at nine stations near Washington, D. C. This is the fresh-water part of the estuary and the mean temperature at each station was between 28.0 and 29.0 degrees centigrade. The mean DO results therefore represent comparable values. While the actual shape of the curve sketched in Figure 15 may be debated, the relationship between mean DO and magnitude of record variance is unmistakable.

Low record variance may be associated with both unpolluted and highly polluted reaches of a river or estuary, while high DO variances would be associated with the midrange of mean values where the processes of deoxygenation and reaeration are competing and about equally effective. The frequency distributions of the records would be skewed toward the midrange of values if the mean values are near zero or saturation, since these represent limiting DO values, while the distribution would probably be Gaussian for means near 50 percent of saturation.

The spectral estimates from records near these limiting values would also be biased, even though the calculation of the spectral statistics would not show it. It is apparent, then, that a scheme of interpretation in which

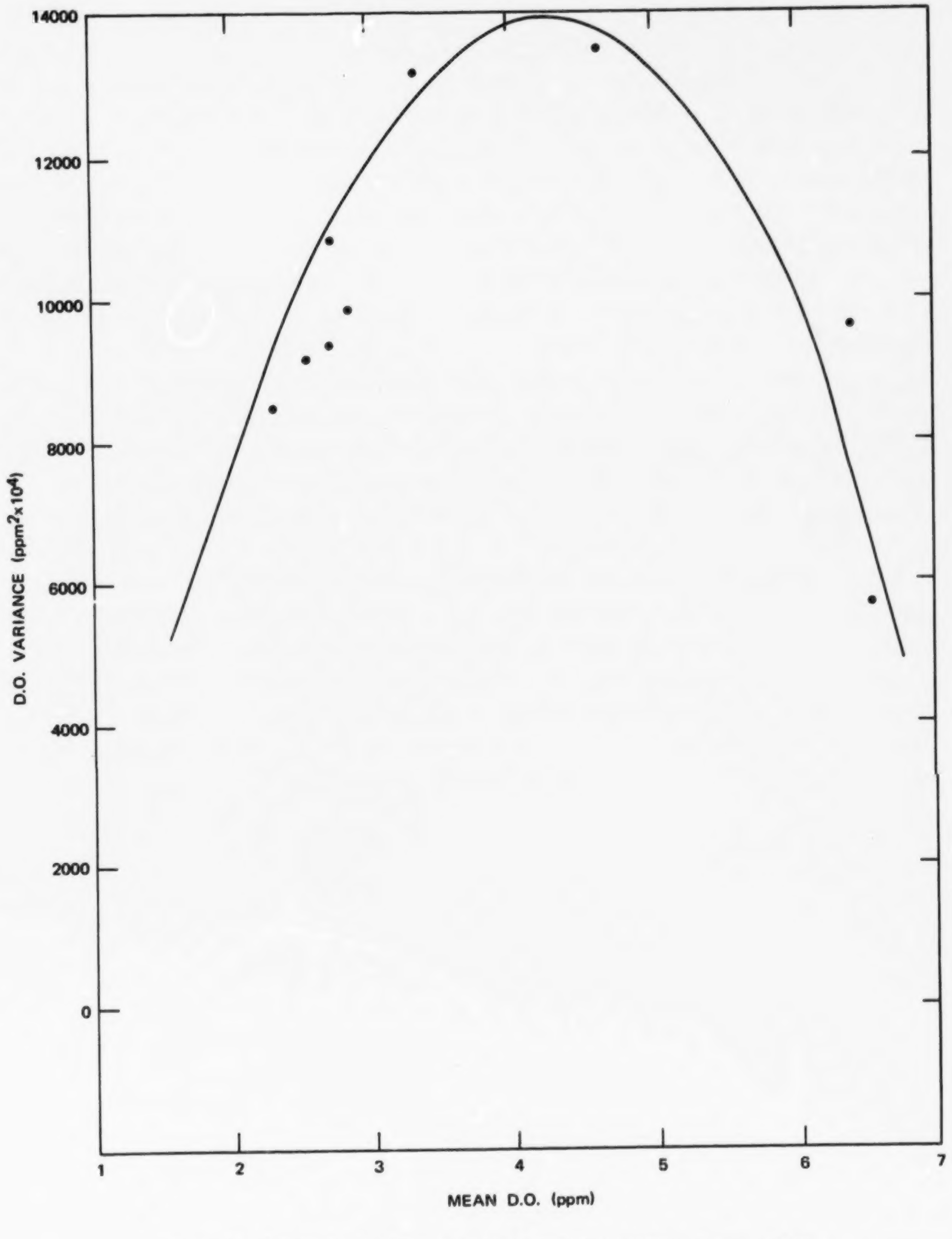


Figure 15. D.O. Variance as a Function of Mean D.O., Potomac Estuary, August. 1959.

the spectral estimates were assumed to result from a sinusoidal fluctuation about the mean (or any other type of uniform fluctuation) could lead to quite erroneous results.

There is one more statistic which may be computed from cross-spectral results and which is peculiarly useful in environmental studies. This is the *response,* which is actually a rate statement of the process between two records. *Response* is a very useful but somewhat dangerous statistic to use, because its dimensions and name imply that a great deal more is known about a process than may actually be the case.

Response is calculated from the response function spectrum and the input spectrum by dividing the spectral estimates of the former by those of the latter and then taking the square root. Another form of estimation of the response is to divide the sum of the estimates of the response function spectrum by the similar sum of the input spectrum and then take the square root. This latter procedure is more logical from an engineering standpoint since what is actually observed in nature is a mixture of the dominant periodic components of both records; this procedure is equivalent to saying that the standard deviations of the response function spectrum and the input spectrum are valid means of quantifying the response of the one record to the other.

This approach is valid providing the input and output records are statistically similar. For example, the frequency distributions of both records should be alike: either both Gaussian, or both skewed in the same direction, or both of similar kurtosis (peakedness). If the two records are not statistically similar, then the use of a response calculation based on them may lead to deceptive results.

The response has units of the output record divided by the input record: ppm DO/ppm BOD, ppm DO/cfs river discharge, ppm BOD/lb sewage load. These rates are, of course, based on variance from the mean values of the input and output records. While they may actually represent fundamental rates for the system being studied, their interpretation should be based on the observed mean values unless there is other contributing information available to support a different interpretation.

8

SPECTRAL ANALYSIS AS A DATA PROCESSING SYSTEM

The major purpose in obtaining and analyzing any set of environmental data is to determine relationships between the parameters measured. This purpose presupposes some vague idea of how the system of interest is operating and what the controlling parameters are. Figure 16 is a generalized flow chart of the dissolved oxygen system in an estuary which shows conceptually how DO may be expected to respond to environmental changes.

Such a general conceptual appreciation of the workings of a system is a prerequisite for any coherent scheme of data processing and analysis. This does not mean that the internal complexities of each part of the system must be known; it does mean that the major interrelationships of the system should be understood.

One of the great advantages of spectral analysis as a scheme of data analysis is that extremely useful quantitative results can be obtained with only the most general knowledge of how the driving forces and the water quality parameters of interest are related.

For example, a record of solar radiation can be analyzed with the DO record for the same period, and information on the effect of sunlight on DO can be obtained without direct knowledge of the photosynthetic oxygen production chain. The results obtained are, of course, purely empirical numbers based on analyses of field data; they do not necessarily represent any new theoretical insight into the system. They are, however, answers about relationships in a particular system at a particular time and, as such, are extremely useful in developing an engineering appraisal of an estuarine water quality problem.

Spectral analysis is most effective and useful when the environmental system under study is dominated by highly variable and complex driving forces. In estuarine systems the complexities introduced by the harmonic nature of tidal and diurnal driving forces make spectral analysis especially useful.

The scheme of spectral analysis as outlined here represents a direct coherent system for the analysis of environmental data. Where records of

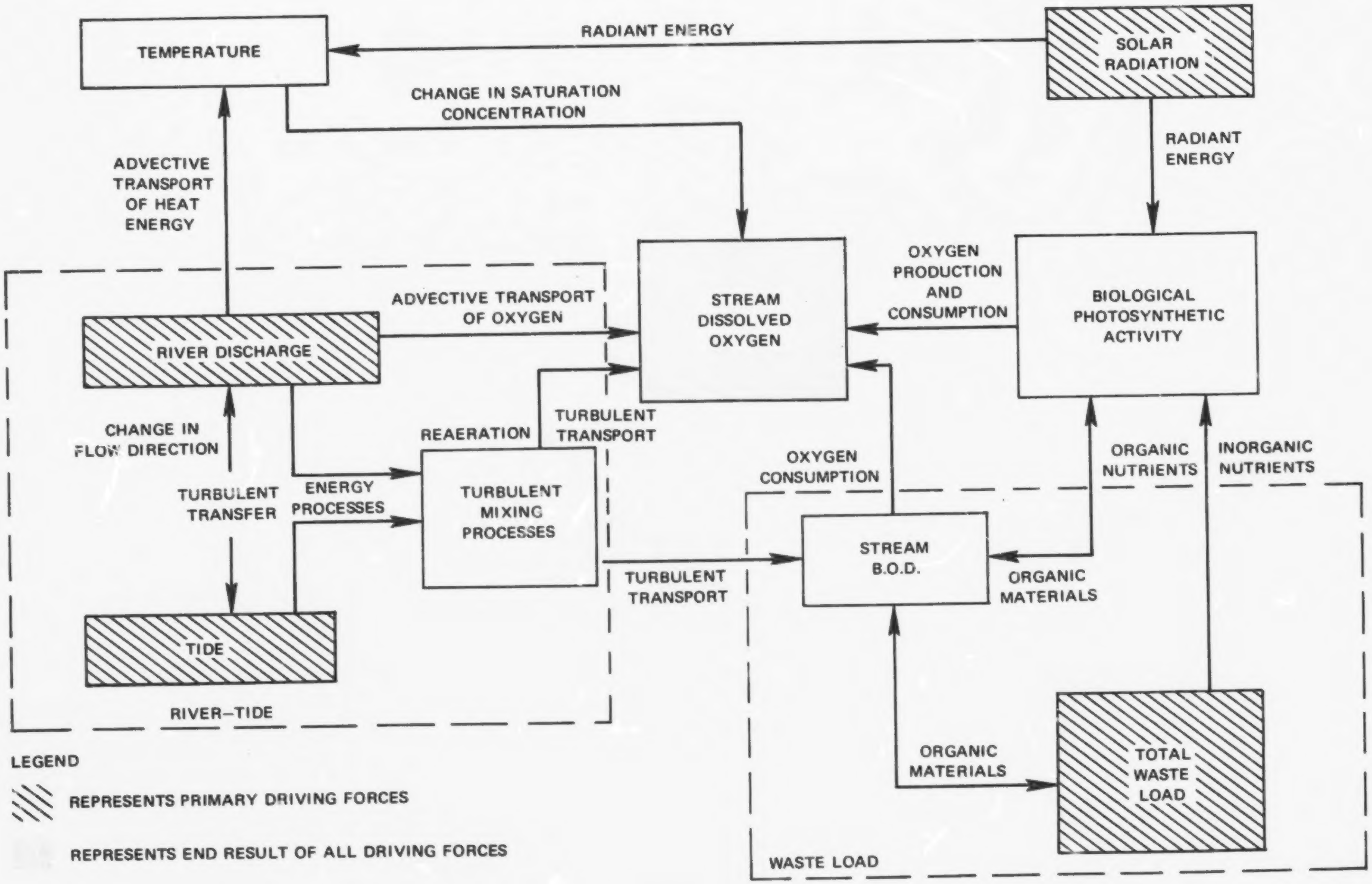

TEMPERATURE
RADIANT ENERGY
SOLAR RADIATION
CHANGE IN SATURATION CONCENTRATION
RADIANT ENERGY
ADVECTIVE TRANSPORT OF HEAT ENERGY
OXYGEN PRODUCTION AND CONSUMPTION
ADVECTIVE TRANSPORT OF OXYGEN
STREAM DISSOLVED OXYGEN
BIOLOGICAL PHOTOSYNTHETIC ACTIVITY
RIVER DISCHARGE
CHANGE IN FLOW DIRECTION
REAERATION
TURBULENT TRANSPORT
OXYGEN CONSUMPTION
ORGANIC NUTRIENTS
INORGANIC NUTRIENTS
TURBULENT TRANSFER
ENERGY PROCESSES
TURBULENT MIXING PROCESSES
STREAM B.O.D.
TURBULENT TRANSPORT
ORGANIC MATERIALS
TIDE
RIVER–TIDE
ORGANIC MATERIALS
TOTAL WASTE LOAD
WASTE LOAD
LEGEND
REPRESENTS PRIMARY DRIVING FORCES
REPRESENTS END RESULT OF ALL DRIVING FORCES

Figure 16. Factors Affecting Stream Dissolved Oxygen Concentration.

data from automatic sampling devices are available, routine processing of the records by spectral analysis would provide nearly all the statistical evaluation of these data that would ever be needed on a routine basis.

With such a scheme of data processing and analysis available, it should be possible to analyze in depth on a routine basis the records from automatic sampling stations. This would reduce by a very large fraction the number of man-hours spent in unprofitable data analysis. A routine data analysis procedure in which basic environmental records, such as temperature, solar radiation, river discharge, and tidal height, were run against water quality parameters such as DO would prove a continuous evaluation of the response of any estuarine system to the environment and establish quantitatively what environmental factors are significant in any particular system.

This is not to say that spectral analysis should be used in all estuarine situations. In cases where manual sampling must be used to measure certain parameters, the situation must be carefully evaluated to determine if the expense of manually collecting data for spectral analysis is justified because the problem at hand is too complex for any other means of analysis to work. This is a matter for individual evaluation.

No scheme of data analysis, particularly one based primarily on statistical manipulation of data, can substitute for engineering knowledge and a logical approach to an environmental problem. Spectral analysis is a scheme of data processing which can produce valid statistical results rapidly and in highly complex situations; the relationship of these statistics to the physical environment must, however, be determined through engineering logic.

This chapter contains five examples of the use of spectral analysis in analyzing and interpreting time series data in river and estuary field studies. The first concerns the tracing of a conservative pollutant in a river, the second concerns the dilution and decay of a nonconservative pollutant in a river, the third illustrates the estimation of diffusion of a nonconservative pollutant in an unstratified tidal river (*i.e.,* negligible salinity), the fourth shows how river flow affects temperature and DO in a tidal system, and the fifth deals with the relationship of river flow to stratification in a stratified estuary.

The design of any spectral analysis computation requires the balancing of several conflicting computational factors; the rationale for choice of design criteria is presented in detail for the first of these examples to illustrate differences in governing criteria for design, and to demonstrate the flexibility of spectral analysis as a statistical tool in extracting different types of information from time series data.

Example 1. *Tracing a Conservative Pollutant in a River*

A discharge of calcium chloride in the wastes from a large chemical plant was affecting the raw water hardness of a municipal water supply some 200 miles downstream. Between these two points was a large reservoir and two large tributary inflows, both of which also had large impoundments.

TABLE 1.a. *Input Record. Salt Waste Discharge ($CaCl_2$ to River 1000 lbs/day)*

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1950	3496	4137	3067	2023	4781	1938	2603	1951	2745	1584	1690	4725
1951	2686	2478	2866	2376	2217	1834	1737	1789	2143	1416	1858	4228
1952	3518	1975	2548	1877	1632	1203	972	993	888	971	2139	1668
1953	1874	2255	2946	1861	3596	1765	2431	1260	1013	809	898	1786
1954	6469	4759	2484	1970	1934	1283	3457	1359	1261	1282	1218	2084
1955	1480	2653	2604	1930	1478	1325	1209	1339	1049	963	1306	1360
1956	1460	1385	1892	2725	2125	1672	1528	1301	1239	1267	1157	2229
1957	2712	3152	1835	1659	1349	1165	1074	808	1272	929	2304	1999
1958	1693	1661	1461	1695	1782	1465	818	1300	1015	936	1172	1769
1959	1788	1298	1984	2165	2115	1288	1195	1556	1051	2250	2979	2048

Information derived from direct sampling of stream water quality, raw water supply hardness, and waste discharges showed that, at a mean flow of 10,200 cubic feet per second over a four-year period, 26.2 percent of the total hardness in the municipal water supply was due to this waste discharge, and from sampling during a subsequent one-year period with a mean flow of 13,600 cubic feet per second, 20 percent of the hardness was due to this waste discharge.

Confirmation of these results was desirable, and cross-spectral analysis of the waste discharge quantities with raw water hardness was used to obtain this confirmation.

The data available consisted of the two time series records shown in Tables 1.a and 1.b. The values in each table are monthly averages of the parameters measured over the same consecutive 10-year period.

Observe that one record (Table 1.a) is in units of thousand pounds/day of calcium chloride while the other (Table 1.b) is in units of mg/1 of hardness. The dimensions carried by records used for cross-spectral analysis need not be identical for computational purposes. Some of the resulting statistics will, however, carry the combined units of the original records; attention should be given, before calculation, to the most useful units for expression of the final results and any necessary conversion made immediately.

In the present example a response in terms of mg/1 hardness per lb/day $CaCl_2$ would result; this is a useful number for the purposes of the analysis, so no conversion is needed and the records can be used as they are. The record length and sampling interval are already established by the data available, so that the only design criterion which can be altered is the number of lags used in computation. Since there are 120 points in each record, the rough 10 percent rule-of-thumb suggests that 12 lags can be used readily. Recent practice in spectral analysis suggests that a much larger number of lags can be used (as much as one-third of the record length has been used by the author, even on very short records, with good results).

TABLE 1.b. *Output Record. Water Supply Hardness (mg/l)*

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1950	102	89	75	74	103	104	106	102	90	102	180	123
1951	132	115	72	91	79	105	87	104	126	141	130	129
1952	95	93	73	76	103	116	109	124	125	125	134	138
1953	118	90	104	102	74	111	126	127	128	141	132	139
1954	87	77	64	68	62	88	122	114	139	128	137	120
1955	120	95	79	86	86	88	97	118	116	119	120	120
1956	130	70	97	87	94	106	96	114	125	132	141	109
1957	117	76	78	80	78	81	79	90	92	93	88	86
1958	94	80	75	75	81	93	93	108	108	109	115	131
1959	95	72	110	93	84	66	94	87	95	88	106	101

To decide the number of lags to use in this example, these factors were considered:

1. What will be the major periodic fluctuations in the system?
   Since the waste discharges are being carried along in the river flow, the major periodic components in hardness variation at the water supply will be closely related to the hydrologic cycle. This cycle has a very strong annual component, and because of the large impoundments on the river and its major tributaries, much of the shorter period fluctuations in flow will be damped out and the annual cycle in flow will govern the downstream hardness variations. It is desirable then, to choose a number of lags in which the annual cycle will be shown clearly.
2. How much resolution is needed?
   High resolution, which is obtained by using a large number of lags, is needed where there may be periodic components of nearly the same frequency in the data. For example, if we suspected that there were significant variations in the data which recurred at 10-month intervals as well as at 12-month intervals, we would go to a high enough number of lags so that these two components would be separated in the analysis (this point will be illustrated subsequently). In the present case the existence of the strong annual cycle and the damping effect of the dams suggests that the annual cycle will dominate the entire spectrum, and that other periodic fluctuations in the flow and in the hardness will be suppressed.
   If there is doubt about what periodic components may be present in any set of data, the best approach is to compute power spectra on the individual records at several different lags, and from these results observe what periodic components are significant; the final design of analysis can then be made with considerable confidence.
3. How much statistical reliability is needed?
   The higher the number of lags used, the wider is the confidence band for each estimate. When a purely random process governs a system, this is a very significant problem; in a system which is

highly deterministic the statistical measures of confidence limits are much less important. Even so, it is good practice to keep the number of lags used in the analysis as low as is compatible with the resolution needed.

The actual choice of what number of lags to use is determined by selecting several possible numbers of lags, calculating the recurrence periods for which these choices will give spectral estimates, if necessary calculating the degrees of freedom of each estimate and the confidence interval, and then making a choice among them.

For the present case 6, 12, 15, and 18 lags all seemed to be equally reasonable. The equation for calculating the recurrence period corresponding to each lag number was given earlier (Equation 6) and is

$$T_r = \frac{2m\Delta t}{r}, \tag{6}$$

where

T_r = period corresponding to the rth lag,
Δt = time interval between data points (1 month here),
m = total number of lags,
r = lag number, $o \leq r \leq m$.

For each of the choices shown above, the recurrence periods at which spectral estimates will be obtained were calculated and are presented in Table 2.

The 6-lag, 12-lag, and 18-lag choices all will provide spectral estimates for a 12-month period. The 15-lag choice will give spectral estimates for 15-month and 10-month recurrence periods, but none for 12 months. This means that a strong 12-month spectral component would be averaged between the 15-month and 10-month estimates if this number of lags is used; this is not desirable, so this choice is rejected.

The 6-lag choice has a 12-month period estimate, but it is adjacent to the long-period (∞) estimate. If either record has a variance component greater than one year in recurrence, then the estimate of this component will overlap the 12-month estimate and it will be impossible to separate them. Therefore the 6-lag choice is ruled out.

The 12-lag choice has one estimate between the long-period and 12-month estimates, and the 18-lag choice has two. The higher resolution of the 18-lag choice is at the cost of some loss of statistical confidence in the result.

The numbers of degrees of freedom for each spectral estimate of the 12-lag choice and the 18-lag choice were calculated using the method given by Blackman and Tukey, which is based on the Chi-square test. This calculation proceeds in this manner:

(1) Calculate effective record length (T_m) from actual record length (T) and total number of lags used (m):

$$T_m = T - 1/3\ (m) = 120 - 1/3\ (12) = 116$$

(2) Calculate equivalent band width (W_e) for each estimate

$$W_e = \frac{1}{m} = \frac{1}{12}$$

TABLE 2. *Recurrence Periods at Various Choices of Total Lags*

Lag Number	Period in Months at Each Lag Number for Total Lags of:			
	6	12	15	18
0	∞	∞	∞	∞
1	12.0	24.0	30.00	36.00
2	6.0	12.0	15.00	18.00
3	4.0	8.0	10.00	12.00
4	3.0	6.0	7.25	9.00
5	2.4	4.8	6.00	7.20
6	2.0	4.0	5.00	6.00
7		3.4	4.30	5.10
8		3.0	3.75	4.50
9		2.7	3.33	4.00
10		2.4	3.00	3.60
11		2.2	2.70	3.30
12		2.0	2.50	3.00
13			2.30	2.80
14			2.10	2.60
15			2.00	2.40
16				2.25
17				2.10
18				2.00

(3) Calculate change in frequency between spectral estimates (Δf)

$$\Delta f = \frac{1}{2T_m} = \frac{1}{(2)\ (116)}$$

(4) Calculate number of degrees of freedom (k) for each estimate

$$k = \frac{W_e}{\Delta f} = \frac{(2)\ (116)}{12} = 19.3$$

This is very roughly equivalent to saying that a person could have as much confidence in the magnitudes of each spectral estimate as he could have in the average value of 20 pieces of data. For the 18-lag computation a similar calculation gave 12.3 degrees of freedom, or a confidence level equivalent to that of having an average based on 13 pieces of data. Since both lag choices have the resolution required, and the numbers of degrees of freedom are both within reasonable working ranges for water pollution engineering, the choice between the two becomes a subjective one.

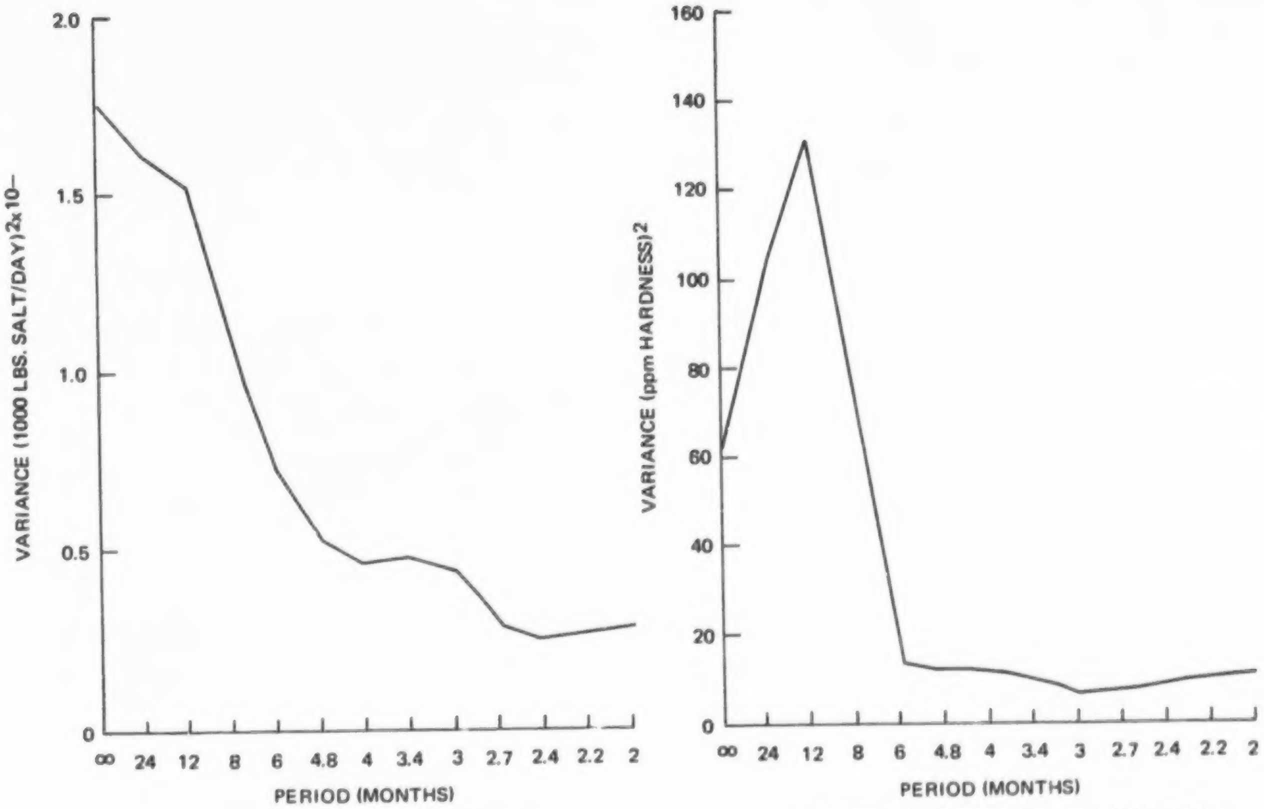
The 12-lag choice was made, primarily because computation time would be considerably shortened while a sufficiently high degree of resolution was maintained.

The cross-spectral analysis computations were then made using the salt waste discharge record as the input and the water supply hardness as the output. Table 3 is a representation of the computer output obtained by cross-spectral analysis of these two records. The computer program was designed so that the output contained not only the individual power spectra and the cross-spectrum of the two records, but also many subsidiary cal-

TABLE 3. *Computer Output of Cross-Spectral Analysis: Salt Waste—Water Hardness Problem*

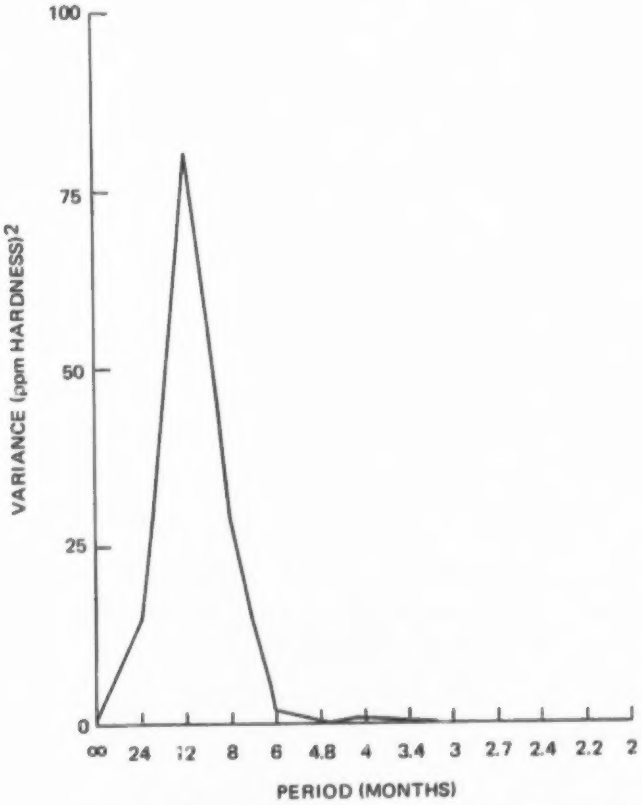
```
                         CROSS-SPECTRAL ANALYSIS
INPUT RECORD — SALT WASTE — 1000 LBS/DAY CACL2
OUTPUT RECORD — WATER SUPPLY — MG/L WATER HARDNESS
120 VALUES              12 LAGS              1.0 TIME INTERVAL

        INDIVIDUAL SPECTRA                CROSS-SPECTRA COMPONENTS
LAG        INPUT         OUTPUT              COSPECT.     QUADSPECT.
 0       1.744E 05     6.322E 01           5.294E 02       0.
 1       1.622E 05     1.013E 02          —9.280E 02     —1.287E 03
 2       1.536E 05     1.322E 02          —2.234E 03     —2.746E 03
 3       1.198E 05     6.999E 01          —9.244E 02     —1.616E 03
 4       7.130E 04     1.383E 01           2.196E 02     —3.057E 02
 5       5.054E 04     1.195E 01           6.219E 01     —3.762E 01
 6       4.434E 04     1.192E 01          —2.477E 01      1.711E 02
 7       4.677E 04     1.008E 01           2.202E 01      1.282E 01
 8       4.009E 04     6.863E 00          —3.384E 01     —1.291E 02
 9       3.177E 04     7.180E 00          —7.806E 01     —8.924E 01
10       2.730E 04     8.875E 00          —2.802E 01     —1.150E 02
11       2.916E 04     9.859E 00          —5.617E 01     —1.010E 02
12       3.267E 04     1.026E 01          —1.314E 02       0.

LAG       PERIOD       COHERENCE           RESPONSE   PHASE (DEG.)
 0         LONG         0.0254             1.606E 00          0
 1        24.00         0.1531             1.551E 01      234.2
 2        12.00         0.6175             8.163E 01      230.9
 3         8.00         0.4135             2.894E 01      240.2
 4         6.00         0.1436             1.986E 00      305.7
 5         4.80         9.0087             1.040E 01      328.8
 6         4.00         0.0566             6.751E 01       98.2
 7         3.43         0.0014             1.402E 02       30.2
 8         3.00         0.0647             4.440E 01      255.3
 9         2.67         0.0616             4.424E 01      228.8
10         2.40         0.0579             5.147E 01      256.3
11         2.18         0.0464             4.571E 01      240.9
12         2.00         0.0515             5.280E 01      180.0

LAG     COHERENCE ROOT
 0            .159
 1            .391
 2            .786
 3            .642
 4            .379
 5            .093
 6            .238
 7            .037
 8            .254
 9            .248
10            .241
11            .215
12            .227
OVERALL RESPONSE = .01162
```

culations which are very useful, but are time consuming to calculate without the use of the computer. Some of these are not needed in the present example, and will be explained when included in the other examples presented here.



A. SPECTRUM OF SALT DISCHARGE

B. SPECTRUM OF WATER PLANT HARDNESS



C. RESPONSE SPECTRUM OF HARDNESS TO SALT DISCHARGE

Figure 17. Spectra from Salt–Hardness Analysis.

Figure 17 shows the individual power spectra of each record, as well as the response spectrum of the pair of records as also presented in Table 3.

The waste discharge spectrum in Figure 17.a shows a pronounced long-period component and a "probably" strong annual component. The word "probably" is used because there could be a strong 24-month component and the annual component would reflect the overlapping estimates of the annual and 24-month components. However, the variance estimates at

8-month and 6-month periods drop rapidly and consistently from the magnitude of the 12-month estimate, so it is probable that the magnitude of the 24-month spectral estimate is the result of overlap between the long-period and 12-month estimates. If what was desired from anlysis of this record was the spectrum of the waste discharge, this analysis would be inadequate and there would be two courses of action open:

First, repeat the analysis at a higher number of lags to ascertain the magnitude of the annual component by computing estimates at several periods between the long-period and annual periods. (See the preceding discussion on the higher resolution obtained by using 18 lags instead of 12 lags.)

Second, detrend the record by running a linear (or second-order) regression of waste discharge against time, compute deviations from the regression line instead of the mean of the entire record, and then compute the spectrum on these deviations. This is a potentially misleading procedure, since the investigator is committed to the assumption that the regression line actually is a deterministic input into the record, instead of a merely statistical device for making something out of nothing.

Fortunately, in the present example, the actual spectrum of the waste discharge is important only as it relates to the hardness of the water supply downstream.

The spectrum of the water supply hardness, Figure 17.b, shows a very pronounced annual component, a smaller amount of long-period variance, and very small variance recurring at other periods. It is apparent that the 12 lags in this case have been quite satisfactory in resolving and quantifying the variance of the water supply hardness record. The same approach used in the further analysis of the waste discharge record would be applicable for this record if additional information were required from the analysis of this one record.

Figure 17.c shows the response spectrum of the water supply hardness record to the waste discharge record. The dominant feature of this spectrum (which results from a calculation using the cross-spectra and coherence of the two records) is the annual component of variance; all other variance components are of negligible magnitude. Such a strong response at one recurrence period and a lack of response at other periods suggests that the assumptions about the importance of the impoundments, in smoothing out all except the annual flow variations, were quite correct; these results also suggest that the water supply hardness is a highly deterministic function of the salt waste discharge.

In Table 3 are shown the square roots of the coherence for each spectral estimate. The annual recurrence period has a coherence root, roughly equivalent to a correlation coefficient, of 0.8, which is a very high result for this type of analysis.* The coherence reflects the amount of correlation between the magnitude of variance in input and output records for each recurrence period, as well as the consistency in the repeating of the period of recurrence itself. In this case, if there were some variation in the time of annual peak runoff from year to year, there would be a decrease in coherence even if the magnitude of hardness variation were the same.

---

* While no detailed analysis of the relationship between coherence root and correlation coefficient has been made as yet, the theory of probability suggests that coherence root is analogous to the correlation coefficient for a joint probability density function, and it would therefore be the product of two correlation coefficients; one relating variance magnitude, and one relating consistency of periods. If this analogy is accepted, then the coherence root of 0.8 would be equivalent to the product of two such correlation coefficients, each of about 0.9.

The results presented in Table 3 contain the information necessary to calculate the amount of hardness in the water supply attributable to the salt waste discharge.

The response spectrum is in the same units as the output spectrum, (*i.e.*, ppm hardness)^2; this spectrum is what the output record spectrum would be if the input record alone generated the output. The *single estimate unit response, i.e.*, ppm hardness per 1000 lb/day salt discharge, is the square root of the ratio of each response spectrum value to each input spectrum value for each of the recurrence period estimates. The *overall unit response* is the square root of the ratio of the summed response spectrum values to the summed input spectrum values and has the same units as the single estimate unit response.

Both of these are valid statistics, each particularly informative in certain types of situations.

The single estimate unit response, R_s, is most useful when there is no transfer of energy from one recurrence period in the input to another recurrence period in the output. For example, if the major recurrence periods in both the salt discharge and water supply hardness records in the present example had been the annual cycle, then the assumption logically would be made that changes in the annual cycle in the input would bring about changes in the annual cycle of the output record, and R_s would be a suitable response statistic to use.

The overall unit response, R_o, is used when there is evidence that there has been a transfer of energy from one recurrence period in the input spectrum to a different recurrence period in the output spectrum. In the present example there has been a transfer of energy from the long-period estimates in the spectrum of the input record to the annual estimate in the spectrum of the output record. This transfer of energy between periods is the result of reservoir storage of salt discharge for many months within the river system, followed by its release from storage when the annual hydrologic cycle brings about a partial flushing of the stored salt discharges, which is a major controlling factor in downstream water hardness. For the present example, then, R_o is the appropriate response statistic to use.

This statistic is given directly in Table 3, the computer printout, as *overall response;* for the present example, $R_o = 0.0116$ ppm hardness/ 1000 lbs. per day of salt.

The average percent of salt hardness at the water plant directly attributable to the salt discharge may be calculated directly from R_o in this manner:

$$\text{mean salt hardness } (\%) = \frac{R_o \times \text{mean salt discharge}}{\text{mean hardness}}. \quad (29)$$

In the somewhat oversimplified "mathematical model" of salt hardness given above, there has been made the implicit assumption that the river discharge has no effect on the water plant hardness. The salt hardness estimates using stream water quality data were, however, based on the observed condition that total hardness varied inversely with river discharge and on the conclusion, from this and other data, that tributary inflow of salt was in direct proportion to river flow and had little effect on total hardness concentration.

The total hardness is, therefore, affected by the varying dilution of the salt waste in the river, an inverse function of river flow.

The cross-spectrum calculation gives results which are averaged for the entire period of record and are related directly to the mean values of all controlling conditions during the period of record. In this case $R_o = 0.012$ is the overall unit response at a mean salt discharge of 1,935,000 lbs/day and a mean water plant hardness of 102 ppm. It also applies for the mean river flow of 12,450 cfs during the period of record. To include the diluting effect of the river flow on the salt waste, the earlier expression for salt hardness is adjusted by including a dilution factor term in the classical form:

$$V_m S_m = V_n S_n, \tag{30}$$

where V_m and V_n are river discharges and S_m and S_n are corresponding salt hardness concentrations.

This can be included in equation (29) to give the result:

$$S_n = S_m \frac{(V_m)}{(V_n)} = \frac{RW}{H} \frac{(V_m)}{(V_n)}, \tag{31}$$

where

W = salt discharge for survey n (lbs/day),
H = mean hardness (ppm),
V_m = mean river discharge (cfs),
S_m = mean salt hardness concentration (ppm),
V_n = river discharge for survey n,
S_n = salt hardness for survey n.

Table 4 summarizes comparable calculations of salt hardness for each of the two field surveys and for the mean of the two surveys.

From the stream water quality data obtained at two different periods, corresponding salt hardness percentages of 26.2 and 20 percent were obtained, an average of 23.1 percent which compares well with the 22.8 percent calculated with equation (31) in this manner:

For the entire period of record . . .

$$S_n = \frac{(.012)(1935)}{102} \frac{(12450)}{(12450)} = 22.8$$

For the two surveys . . .

$$S_n = \frac{(.012)(1852)(12450)}{106(10240)} = 25.8$$

$$S_n = \frac{(.012)(1722)(12450)}{98(13600)} = 19.3$$

All of the differences between corresponding salt hardness values shown in Table 4 are within the limits of precision of the chemical determination for hardness.

The cross-spectrum calculation thus provides an excellent check of the other calculations. The spectral analysis is based only on the salt waste discharge and the water plant hardness, not on the stream survey data, so the two methods of calculation are independent.

The cross-spectral analysis of these two records provides another statistic not obtainable by the other means of calculation used in the analysis of these data. This is the mean time of passage of the salt waste from the point of discharge to the water plant.

Table 3 shows the phase lags associated with each recurrence period. The annual recurrence period is a major feature in the spectrum of the output record and also has the highest coherence between records; this annual cycle reflects the yearly flushing of the impoundments during high spring river flows. The 7.7-month phase lag shown in Table 3 for this annual cycle is an average time of passage of salt waste between the two points of record; in this case this number probably represents reservoir residence time to a large extent.

Knowledge of time of passage can be used as a pollution control measure through timing of releases to minimize effects at downstream points.

This example shows how information in available plant records can be extracted through use of spectral analysis. In the present case not only did the analysis of plant records alone provide the same information, which, using conventional techniques, required a stream survey in addition to the plant records, but it also provided additional useful information.

A slightly more refined approach to this particular problem would include the cross-spectral analysis of river discharge and water plant hardness records to compute an overall unit response factor for hardness to river discharge; this factor would be a more precise measure of the relationship of hardness to river discharge than the dilution effect only, as assumed here.

Example 2. *Dilution and Decay of a Nonconservative Pollutant in a River*

The first example showed how the effects of a conservative pollutant can be examined quantitatively using spectral analysis. In the present example, a similar approach is used in the evaluation of the effects of two sewage discharges on a river. This example is based on the standard sanitary engineering concepts of the relationship of DO and BOD, with the additional use of the sunlight-DO photosynthesis relationship exhibited in Figure 16 and quantified through the cross-spectrum calculations.

In this example, the starting point is the waste load of the sewage treatment plants as described by their BOD loads; the impact of these loads on the stream is traced through the changes in stream BOD and DO as they vary with waste loads, with sunlight, and with river flows.

The physical situation is that shown in Figure 18. The two waste outfalls were sampled at three-hour intervals to obtain 127 consecutive

6

TABLE 4. *Comparison of Salt Hardness Values*

Data Base	Mean Salt Discharge (lbs/day)	Mean Hardness (ppm)	Mean River Discharge (cfs)	Based on Water Quality Data, Salt Discharge, Water Plant Hardness	Based on Spectral Analysis of Salt Discharge and Water Plant Hardness
Ten Years of Record of Salt Discharge and Water Plant Hardness	1,935,000	102	12,450	23.1	22.8
Four Years of Record of Salt Discharge, Water Plant Hardness, and Stream Water Quality Data	1,852,000	106	10,240	26.2	25.8
One Year of Record of Salt Discharge, Water Plant Hardness, and Stream Water Quality Data	1,722,000	98	13,600	20	19.3

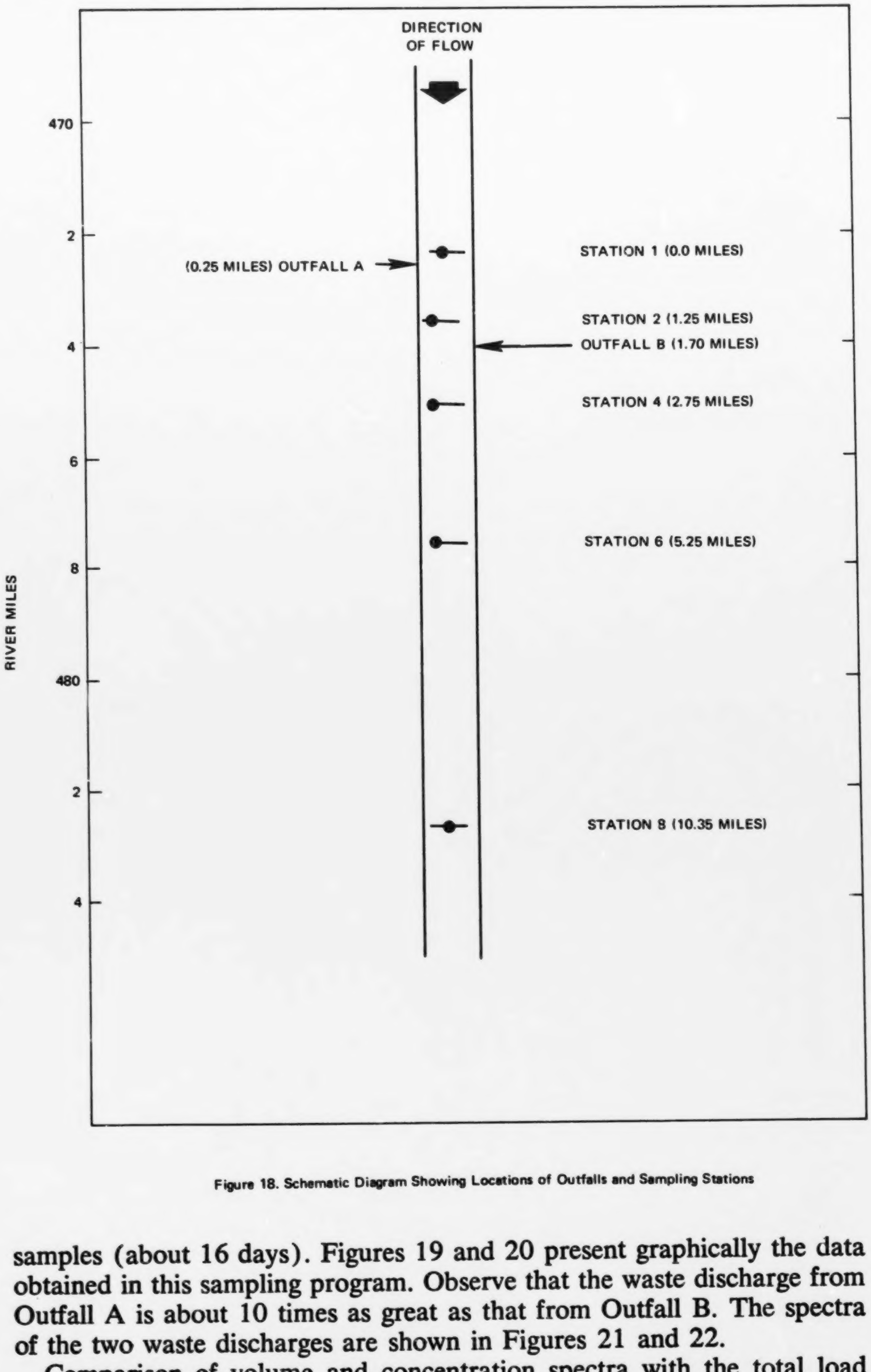

DIRECTION OF FLOW
RIVER MILES
470
2
4
6
8
480
2
4
(0.25 MILES) OUTFALL A
STATION 1 (0.0 MILES)
STATION 2 (1.25 MILES)
OUTFALL B (1.70 MILES)
STATION 4 (2.75 MILES)
STATION 6 (5.25 MILES)
STATION 8 (10.35 MILES)

Figure 18. Schematic Diagram Showing Locations of Outfalls and Sampling Stations

samples (about 16 days). Figures 19 and 20 present graphically the data obtained in this sampling program. Observe that the waste discharge from Outfall A is about 10 times as great as that from Outfall B. The spectra of the two waste discharges are shown in Figures 21 and 22.

Comparison of volume and concentration spectra with the total load spectra from each of these two outfalls shows that the total load of BOD discharges is not closely related to the pattern of volume or concentration. This is of interest because of the tendency to base estimates of the fluc-

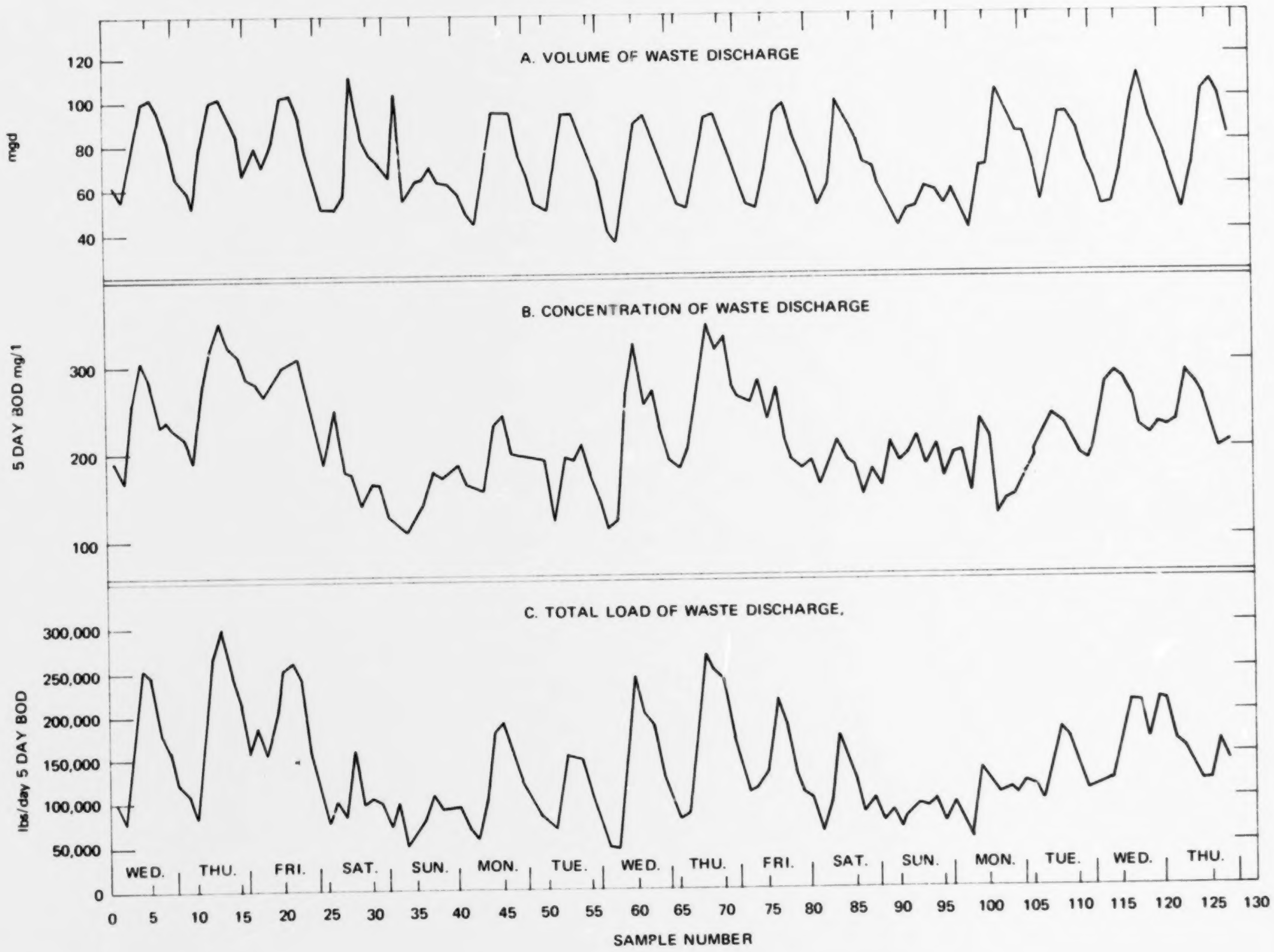


Figure 19. Waste Discharge From Outfall A, Individual Samples.

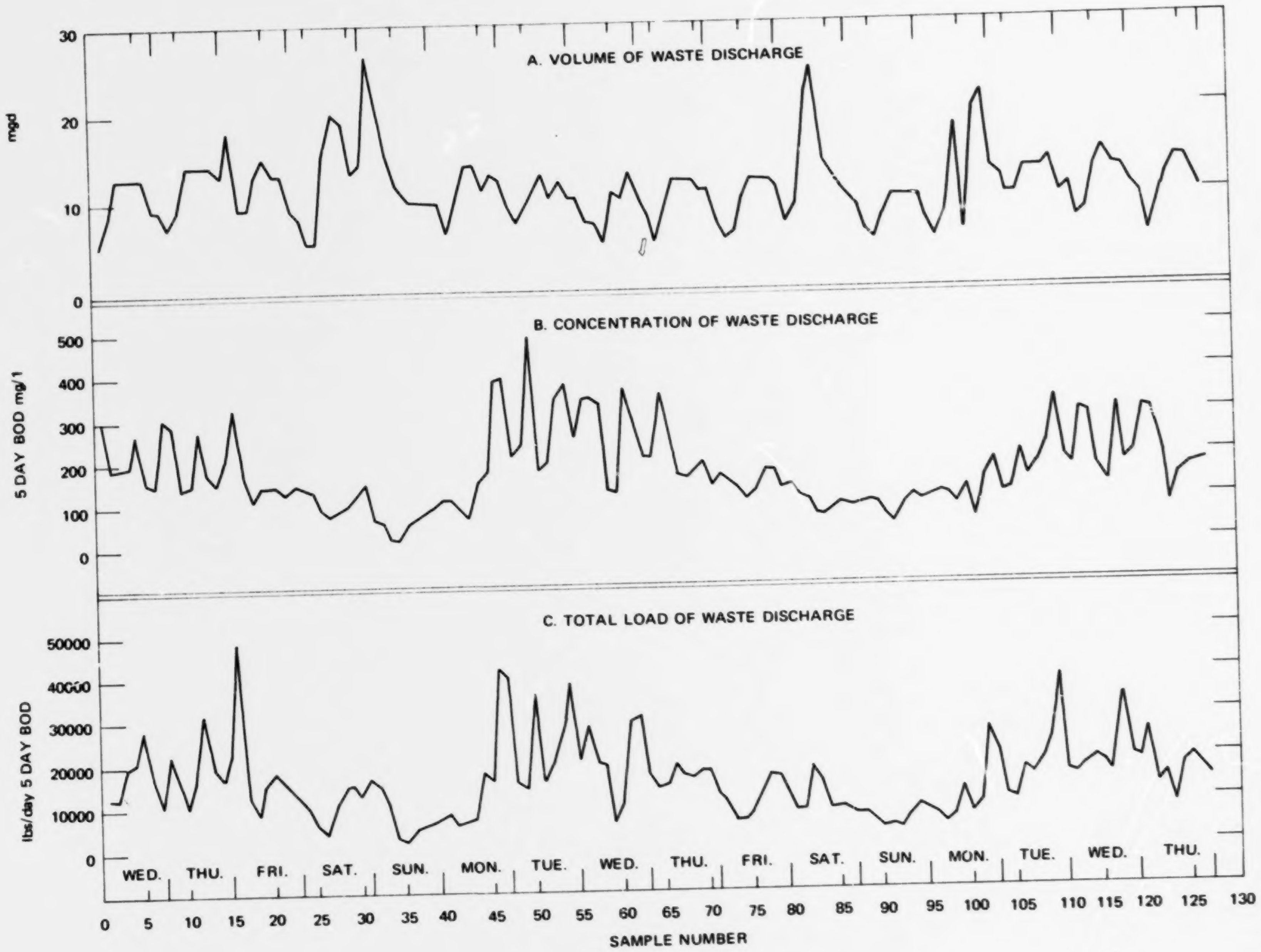


Figure 20. Waste Discharge from Outfall B, Individual Samples.

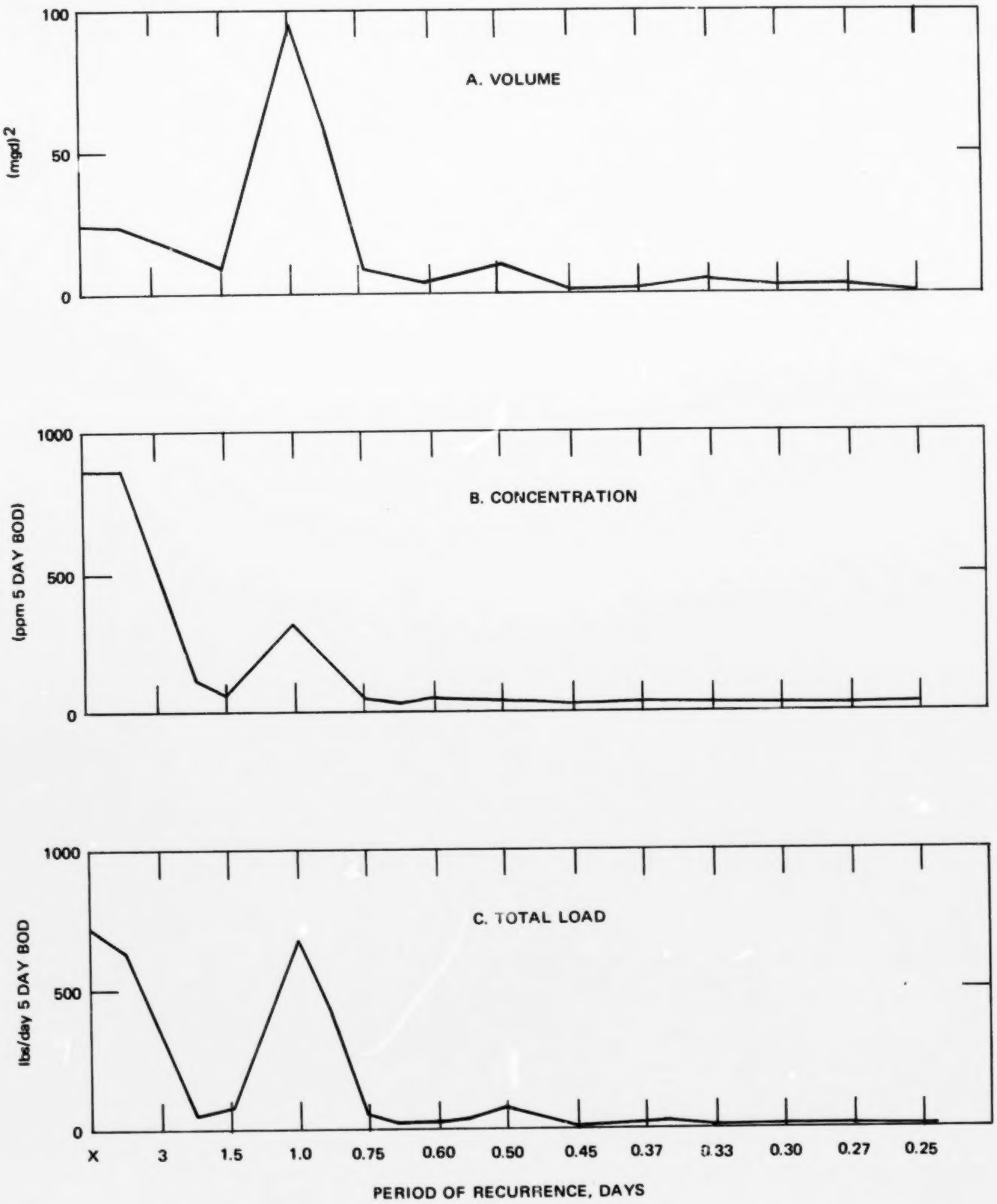


Figure 21. Spectra of Waste Discharge Parameters, Outfall A.

tuation in a sewage plant wasteload on the fluctuations in volume. Total loads are used in the further computations here.

The total load spectrum of Outfall A exhibits long-period and diurnal components of about the same magnitude, while there is also a very small semidiurnal component. If it is assumed that all the periodic fluctuations are sinusoidal (the dangers of this assumption have been discussed), the amplitude of each of these components can be estimated from the spectral estimates.* The sine-wave amplitudes (SWA) corresponding to these

---

* The variance of a sine wave from its mean value is equal to $\frac{1}{\sqrt{2}}$ times the square of its amplitude.

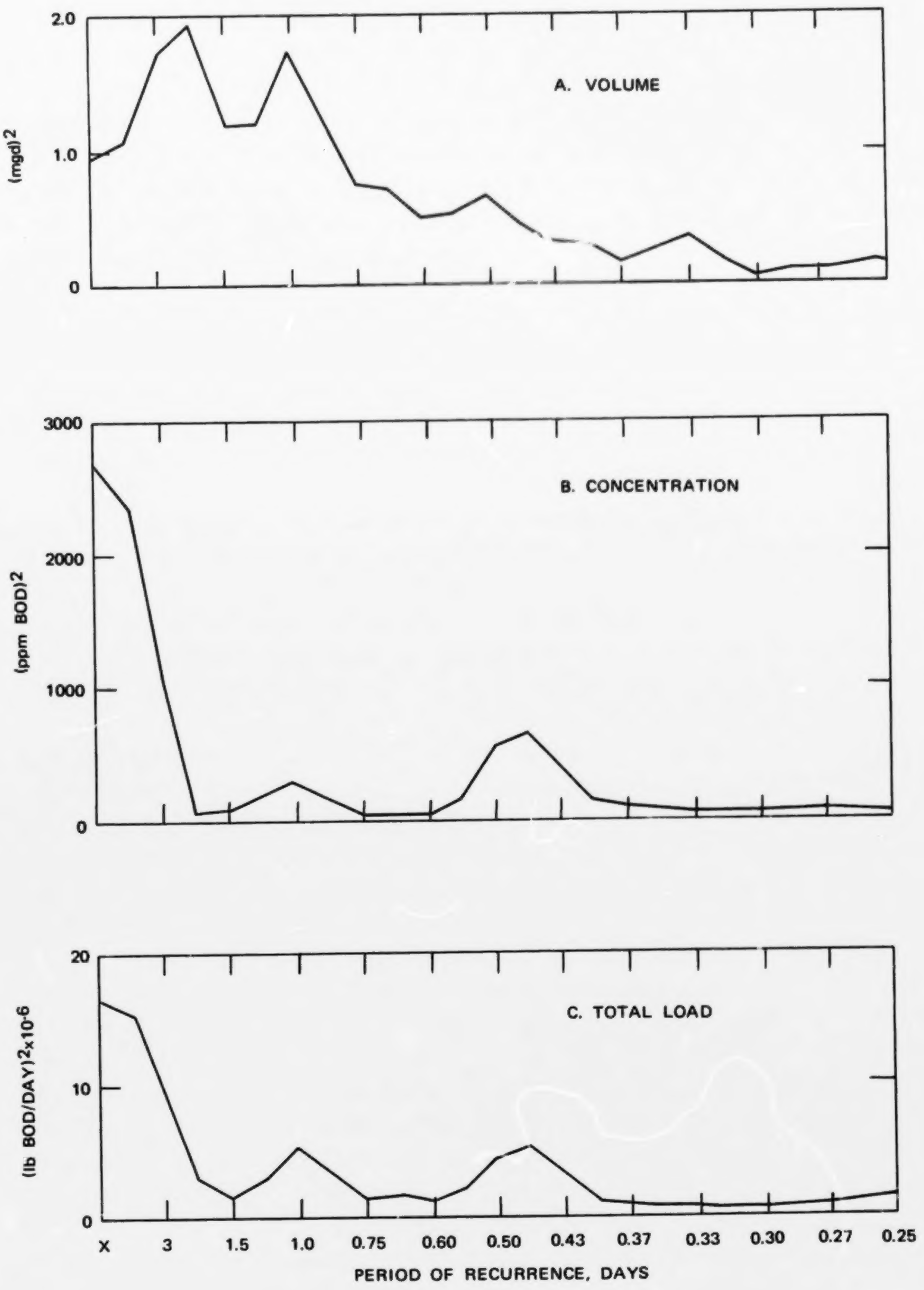


Figure 22. Spectra of Waste Discharge Parameters, Outfall B.

variances are: long-period, 58,000 lbs. BOD_5/day; diurnal, 53,000 lbs. BOD_5/day; semidiurnal, 16,500 lbs. BOD_5/day. The mean was 136,000 lbs. BOD_5/day. The three sets of components indicated account for about 91 percent of the total variance of the record.

The total BOD_5 load spectrum of Outfall B exhibits a large long-period component, SWA 4700 lbs./day, and semidiurnal components, SWA 5500 lbs./day. These account for about 79 percent of the total record variance. The mean waste discharge is 15,100 lbs./day.

Both discharges exhibit wide fluctuations in total load, and the fluctuations in the Outfall B load are proportionately greater than those in the Outfall A load. Outfall A discharges through a diffuser across the river channel, while Outfall B is a point source on one shore. These two facts are of importance in discussing the character of the stream BOD responses to these two outfalls.

Figure 23 presents the spectra of stream BOD and the total load responses of stream BOD for both outfalls at five stream stations. Station 1, which is above both outfalls, exhibits primarily long-period components, and only quite small shorter-period effects. From Station 2 downstream the effects of the strong diurnal components in the waste discharges are quite apparent.

Several environmental conditions, in addition to the spectrum of the entering waste load, determine the BOD response spectrum at any stream station.

First, the relative magnitudes of the BOD load already in the stream and the BOD of the entering waste load can determine whether the BOD response of the stream is large enough to be observed and what is the magnitude of the overall response.

Second, the turbulent-mixing regime and the initial dispersion of the waste in the stream exert some control over the response spectrum. Turbulent mixing in the stream tends to eliminate small variations in concentration first and larger variations later; ultimately responses from all sizes of variations become smoothed out into secular trends in the data. An intense turbulent-mixing regime would tend to bring about this transition more rapidly than a weak one.

Third, the nonconservative nature of the BOD parameter means that there is a time-dependent decreasing trend in the BOD response at stream stations.

Fourth, there may be some indirect response of the stream environment to the waste load, and this may exert some effect on the stream BOD response. For example, nutrients entering stream in the waste load may stimulate algal growth; the increased algal population is then measured as part of the stream BOD at downstream stations.

In the present situation, the discharge from Outfall A represents about 22 percent of the total load in the stream below the outfall, whereas the discharge from Outfall B represents about 2.4 percent of the total load in the stream below that outfall. The response spectra to the two discharges are, however, very similar in magnitude at Station 4 as well as at Stations 6 and 8. In Figure 22 the diurnal response components to Outfall A are also about the same at Station 2 as at Station 4.

The lack of change in the magnitudes of the diurnal components of the response spectra indicates that the longitudinal turbulent-mixing regime in this stretch of the river is not strong enough to affect the initial dispersion of wastes from this outfall. There is a similarity in magnitude of spectral components in response to Outfalls A and B at Station 4. A well-mixed input from Outfall A and a poorly mixed input from Outfall B

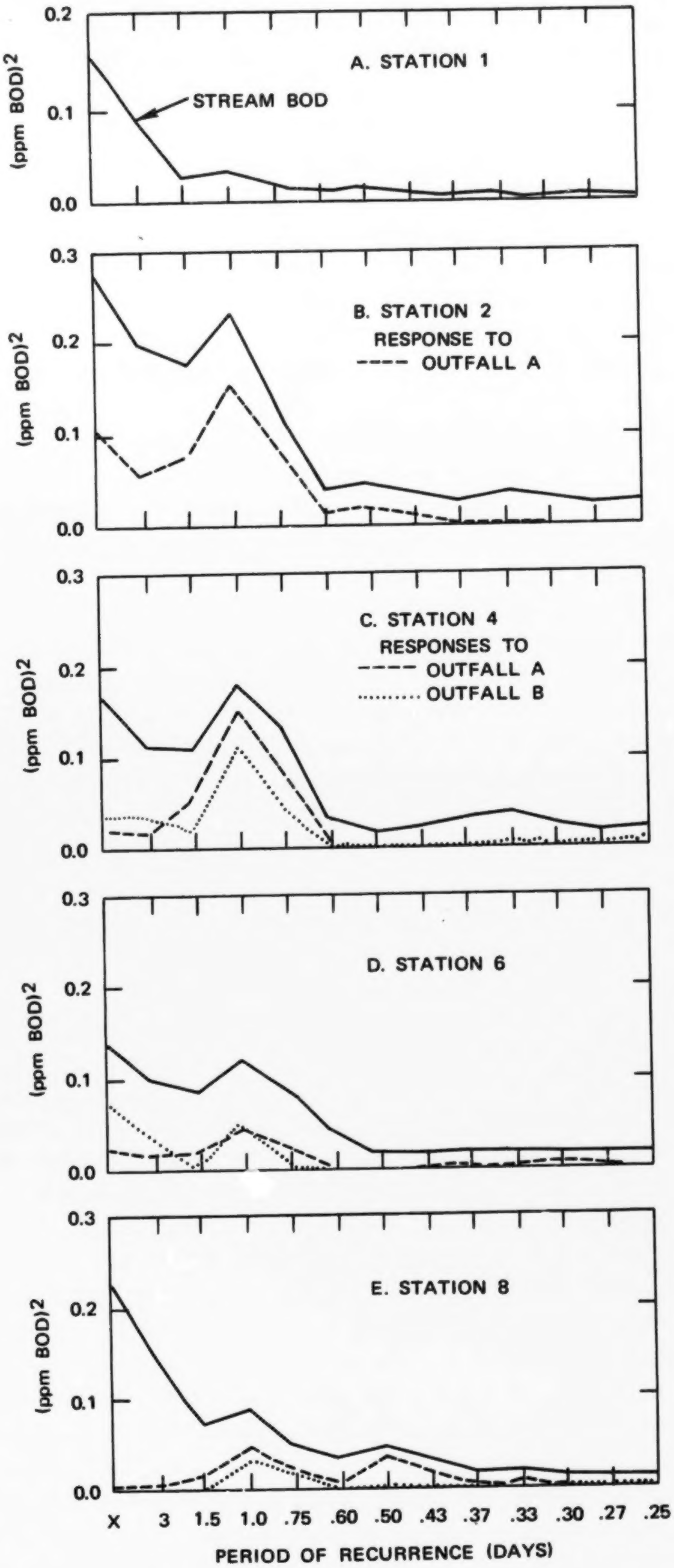


Figure 23. Spectra of Stream BOD with BOD Response Spectra.

affects the stream BOD at Station 4, and it is logical that the relative magnitudes of the response spectra at this station should bear little resemblance to the relative magnitudes of the waste discharges.

The change in mean value brought about by the waste inflow, any secular trends in the waste discharge, the sampling and determination errors in the BOD test, and the longitudinal turbulent-mixing regime govern the long-period response components. Turbulent-mixing phenomena tend to cause a shift in the dominant response components toward the longer periods and ultimately these effects appear as a secular trend in the mean value at a particular sampling station. When a nonconservative parameter, such as BOD, is the criterion considered, the biochemical changes occurring at the time may obscure the shift in the spectral response.

The long-period response components at Station 2 probably reflect secular trends in the discharge for Outfall A, whereas the smaller ones at Station 4 reflect the decreased importance of the long-period fluctuations in waste discharge from both outfalls in the presence of the higher mean stream BOD at this station.

The response spectra at Station 6 show significant changes from those at Station 4. Suppression of the diurnal response components to both outfalls suggests that the controlling phenomenon is the turbulent mixing between Stations 4 and 6, whereas the increase in magnitude of the long-period response components to Outfall B indicates that there has been some shift in response to longer periods as the waste discharge has become mixed into the stream. These may be related to the difference in initial dispersion of the two waste discharges, since the response to Outfall A shows no similar effect.

At Station 8, the diurnal response components to both outfalls are of about the same magnitude as those computed at Station 6, indicating either a somewhat-less-intense turbulent-mixing regime in this section of the river than was indicated in the reach between Stations 4 and 6 or a limit to the effectiveness of turbulent-mixing processes as far as longitudinal waste dispersion is concerned. An interesting phenomenon in the response spectrum to Outfall A is the appearance of a semidiurnal response component. This component does not appear at Stations 4 and 6, and its appearance so far downstream suggests that it may be an indirect result of the waste discharge rather than a direct one. Inorganic nutrients introduced in the waste discharge, that tend to increase the algal population, may explain this component. If this effect does exist, it may affect the entire spectrum at this station.

The rationale presented in the preceding paragraphs serves as the physical basis for a quantitative interpretation of the results of the cross-spectral analysis of these data. The total response, *i.e.,* the sum of all the periodic response components, decreases as stations farther from each outfall are reached. This decrease is a function of the longitudinal turbulent-mixing processes, which tend to smooth out concentration fluctuations resulting from the waste discharge, and of the rate of biochemical oxidation of the total waste load in the stream. It is not possible with present knowledge to separate these two effects in the response spectra to obtain independent estimates of biochemical degradation and turbu-

TABLE 5. *Correlation Coefficients of Response Spectra to Outfall Discharges*

Recurrence period of spectral component, days	Correlation coefficient of response spectrum						
	to Outfall *A* at Station				to Outfall *B* at Station		
	2	4	6	8	4	6	8
x	.61	.37	.40	.16	.48	.72	.18
3	.55	.38	.37	.15	.55	.59	.12
1.5	.66	.73	.49	.52	.43	.28	.23
1.0	.81	.88	.60	.71	.80	.57	.62
0.75	.80	.82	.48	.64	.61	.31	.57
0.60	.67	.42	.13	.44	.13	.35	.25
0.50	.66	.33	.39	.80	.19	.46	.12
0.43	.54	.19	.41	.71	.21	.40	.11
0.38	.36	.14	.39	.35	.30	.25	.19
0.33	.37	.15	.37	.43	.42	.16	.36
0.30	.16	.23	.58	.21	.48	.12	.29
0.27	.13	.18	.38	.26	.49	.08	.22
0.25	.05	.40	.15	.30	.69	.10	.27

Correlation coefficient $\equiv \sqrt{\text{coherence}}$.

lent mixing from the spectral results alone. The spectral results do, however, provide phase lag relationships that can be interpreted as time-of-passage information, and biochemical degradation rate constants have been determined from long-period BOD's on this system. From this information the decrease in response due to biochemical loss in the waste load can be determined, and then some estimate can be made about turbulent mixing and dispersion in this reach of the river.

Phase lags can be calculated for each periodic component of the response spectrum as shown in Example 1. The best phase lag to use for interpretation as a time of passage is obtained from the response component that has the highest degree of correlation with the waste discharge spectrum. Table 5 shows the correlation coefficients computed for each spectral response component at each station for both outfalls. While these correlation coefficients cover a wide range of values, those relating the diurnal responses are fairly high, above 0.8 for Stations 2 and 4. The somewhat lower correlation coefficients at Stations 6 and 8 might be expected since mixing of the waste into the river, which may be characterized as a random process, leads to distortion and ultimate disappearance of the short-period fluctuations related to the waste discharge.

Table 6 presents the phase lag in hours of the diurnal components of stream BOD, as well as responses calculated from spectral results.

The phase lags show the time elapsed between passage of the diurnal peak past the outfall sampling point and its passage past the stream sampling station. Time of passage in the river itself may be obtained by subtracting phase lags to each successive sampling station. The negative phase lag indicated for Outfall B at Station 4 is an artifact of the computation resulting from the fact that the diurnal peak in BOD from Outfall B does not coincide with the diurnal peak in BOD at Station 4; it has no

TABLE 6. *Phase Lags and Responses to Outfall Discharges*

	From Response to Outfall A at Station				From Response to Outfall B at Station		
	2	4	6	8	4	6	8
Phase Lag, hrs. of Diurnal Components	1.4	3.8	6.2	9.9	−1.9	1.3	4.8
Response, ppm BOD/lb. of BOD/day, of Diurnal Components	1.45×10^{-5}	1.39×10^{-5}	0.767×10^{-5}	0.762×10^{-5}	13.6×10^{-5}	8.52×10^{-5}	7.24×10^{-5}
Total Response ppm BOD	.733	.589	.379	.392 .326	.538	.424	.264
ppm BOD/lb. of BOD/day,*		$.959\times10^{-5}$	$.617\times10^{-5}$	$.638\times10^{-5}$	5.65×10^{-5}	4.45×10^{-5}	2.77×10^{-5}

* Without semidiurnal response included.

TABLE 7. *Times of Passage Calculated from Diurnal Response Components*

Reach of River	Calculated from Response to Outfall A		Calculated from Response to Outfall B	
	Time of Passage, hrs.	Velocity MPH	Time of Passage, hrs.	Velocity MPH
Station 2 — Station 4	2.4	0.63	—	—
Station 4 — Station 6	2.4	1.0	3.2	0.78
Station 6 — Station 8	3.7	1.4	3.5	1.5
Station 4 — Station 8	6.1	1.2	6.7	1.1

physical significance in the interpretation of phase lags as times of passage in the river. Table 7 presents time of passage calculated from these phase lags. The agreement between times of passage calculated from each outfall is fairly good. Velocities for this reach of the river in these discharge ranges have been estimated at about 1.0 miles per hour from volume displacement calculations; this is also a good check of the spectral results.

The deoxygenation rate constant for BOD in this reach of the river is assumed to be 0.16 day^{-1} at 20°C for the purpose of estimating the loss in BOD due to biochemical oxidation. At the mean temperature of the water during this study, the deoxygenation rate constant would be about 0.24 day^{-1}. From this rate constant and the time-of-passage estimates of the preceding example, the total response at each station in the absence of any loss through turbulent mixing can be calculated. A measure of the turbulent mixing can then be obtained from the remaining change in total response at each station. The pertinent statistics incorporating this approach are in Table 8.

The statistic of particular interest is the dilution factor, *D*, defined as the ratio of total response (when the BOD loss is due to biochemical loss only) to the observed total response. If it is assumed that each sampling point is representative of a volume of the water passing that point, then the value of *D* is regarded as a volume-of-dispersion ratio. For example, if a unit quantity of the waste discharge from Outfall A is dispersed in a volume *V* as it passes Station 2, then, as this same quantity of waste passes Station 4 it is dispersed in a volume equal to 1.18V. The similarity of values of *D* for Outfall A at Stations 4 and 6 to those for Outfall B at Stations 6 and 8 is very striking. This similarity suggests that the turbulent diffusion characteristics of waste dispersal in this river are more closely related to the concentration of the waste at specific points rather than to some intrinsic property of the receiving stream.

TABLE 8. *Dispersion of Wastes Between River Stations*

Station Number	Waste from Outfall A			Waste from Outfall B		
	R_T	R_L	D	R_T	R_L	D
2	.733	—	—	—	—	—
4	.589	.693	1.18	.538	—	—
6	.379	.557	1.47	.424	.500	1.18
8	.326	.348	1.07	.264	.264	1.47

R_T = total response (BOD mg/1)
R_L = total computed response for BOD loss due to decay only
D = dilution factor = $\frac{R_L}{R_T}$

Although this form of interpretation and calculation of turbulent diffusion processes does not have the intrinsic beauty found in calculations of eddy diffusion coefficients, it does have the virtue of providing an insight into how the waste material disperses in the stream, thereby giving an appreciation of how the waste assimilation capacity of a stream can be utilized most effectively to dispose of waste effluents. It is worthwhile noting that, as shown in Table 8, turbulent diffusion is more significant in reducing the BOD observed at successive stations than are the natural assimilation processes, which are the usual bases for estimating a stream's waste assimilation capacity.

Spectral calculations offer a fresh approach to the problems of turbulent diffusion and waste dispersion in streams. This brief discussion aims to show only how these problems yield to the techniques of spectral analysis. No attempt is made to explore these possibilities fully.

Example 3. *Diffusion of a Nonconservative Pollutant in an Unstratified Tidal River*

The two preceding examples showed the use of cross-spectral analysis in rivers and streams. In these cases the dominant harmonic components were the diurnal variations in sewage discharges and in the photosynthetic production of oxygen.

The present example introduces a major harmonic component common to all estuarine systems—the motion associated with the tide. The situation presented involves sewage discharges entering an unstratified tidal river, so that the effect of tidal movement can be examined free of the complications of salinity-dominated stratification.

This example is an extension of the analysis of the Potomac Estuary near Washington used in Chapter 3 to illustrate the interpretation of individual power spectra. The present discussion is based on the cross-spectra of these BOD and DO records and the use of the solar radiation record in conjunction with them.

The spectrum of solar radiation obtained during the field study is shown as Figure 24. The strong diurnal component is expected, but it is interesting to note the appearance of a significant semidiurnal component.

The source of this component is not known, but several possibilities exist. These data were taken at a Weather Bureau research station and precautions were taken to avoid back-radiation from local night lights or other local environmental artifacts. The semidiurnal component may be a moonlight effect or reflection of night lights of the city from clouds. The means of analysis would regard a situation in which there was a daytime peak of radiation (sunlight) and a nighttime peak of radiation (moonlight or city lights) as a semidiurnal effect of small magnitude. Whatever the cause, it still represents an impingement of energy on the instrument and consequently on the water.

This discussion centers around the cross-spectral analysis of three types of records: solar radiation at one point in the area, DO at nine river stations, and BOD at nine river stations. The cross-spectra computed with these records were:

Input	*Output*
Solar Radiation	BOD
Solar Radiation	DO
BOD	DO

Solar radiation and the effects of waste load (including diffusion and

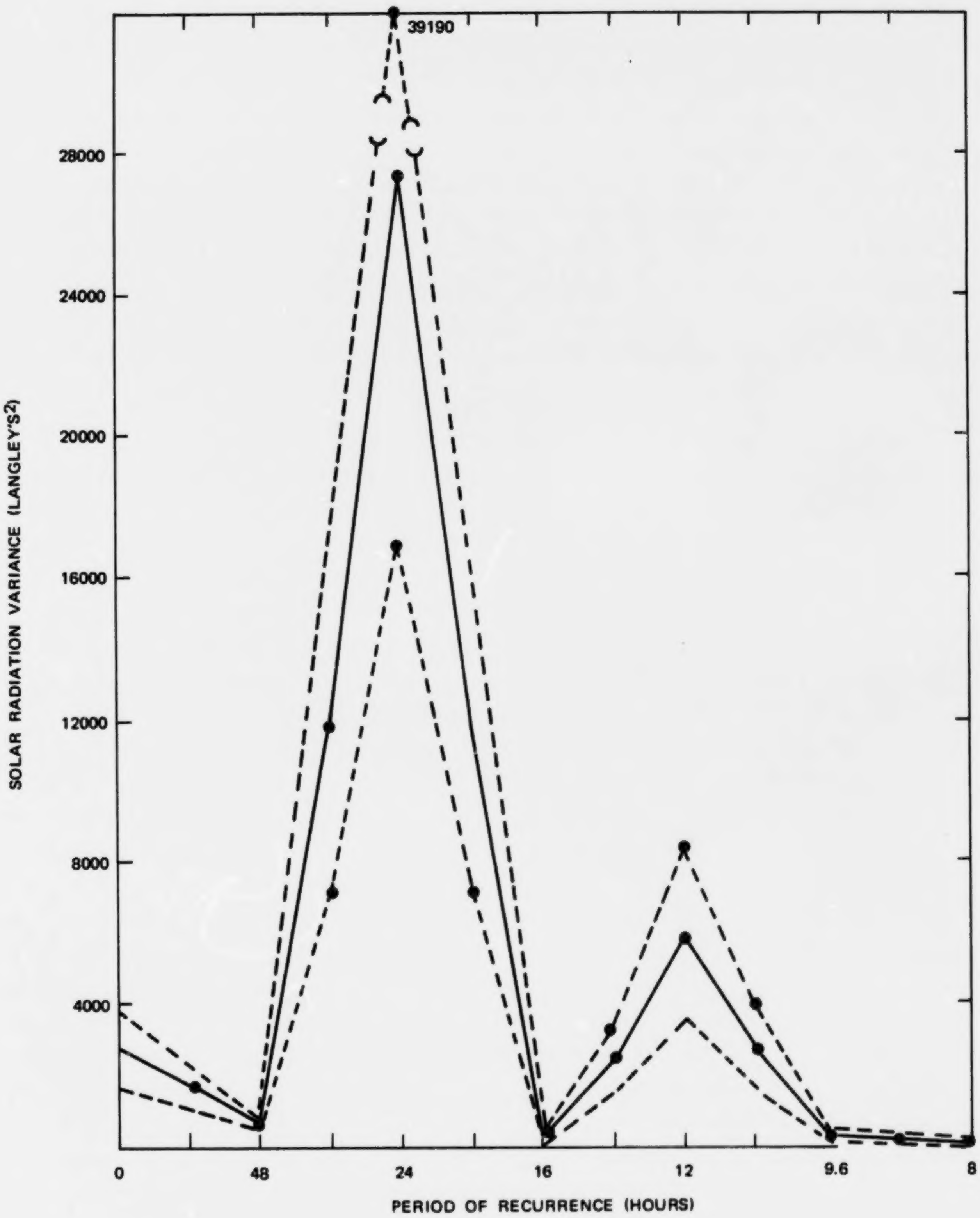


Figure 24. Spectrum of Solar Radiation.

advection of wastes) are the primary driving forces affecting the DO in this system. While the stream BOD is a measure of some of the effects of the waste load, it is also a measure of other things. The algae population, which produces and consumes oxygen, also contributes organic matter to the stream BOD. The DO-solar radiation cross-spectra will show the net effect of solar radiation on DO, which would include both the direct oxygen production or consumption and the change in BOD due to changes in algal cell material, *i.e.,* decayable organic matter.

The individual power spectra of the stream DO and stream BOD at the nine stations on the Potomac estuary are shown in Figures 25 and 26. A

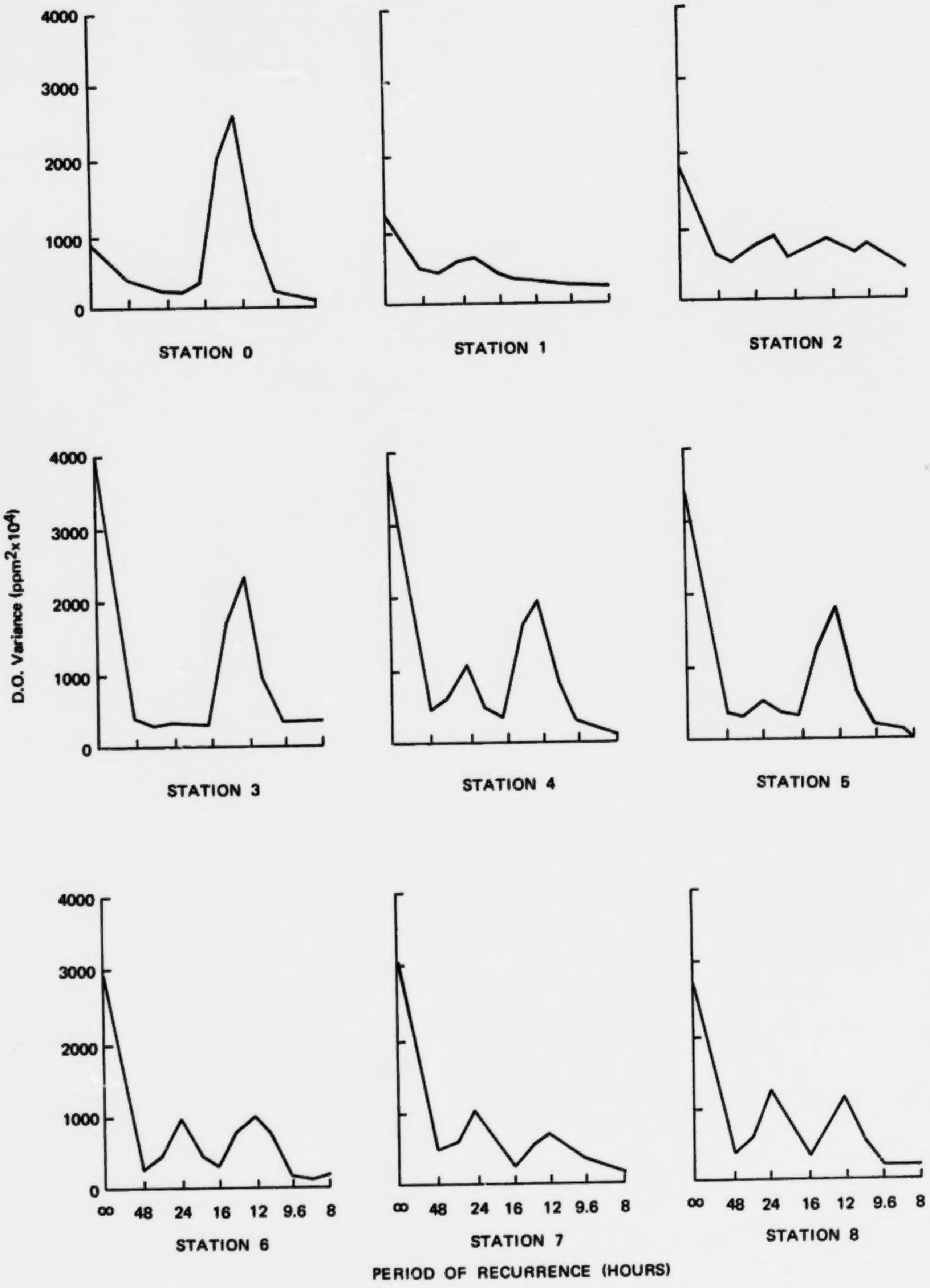


Figure 25. Stream D.O. Spectra in the Potomac Estuary.

schematic representation of the sampling station locations was presented in Figure 11.

The DO response at each stream station to solar radiation is shown in Figure 27. The increasing importance proceeding downstream of diurnal fluctuations and longer variations is quite obvious. The suppression of semidiurnal response in comparison to the strong semidiurnal components

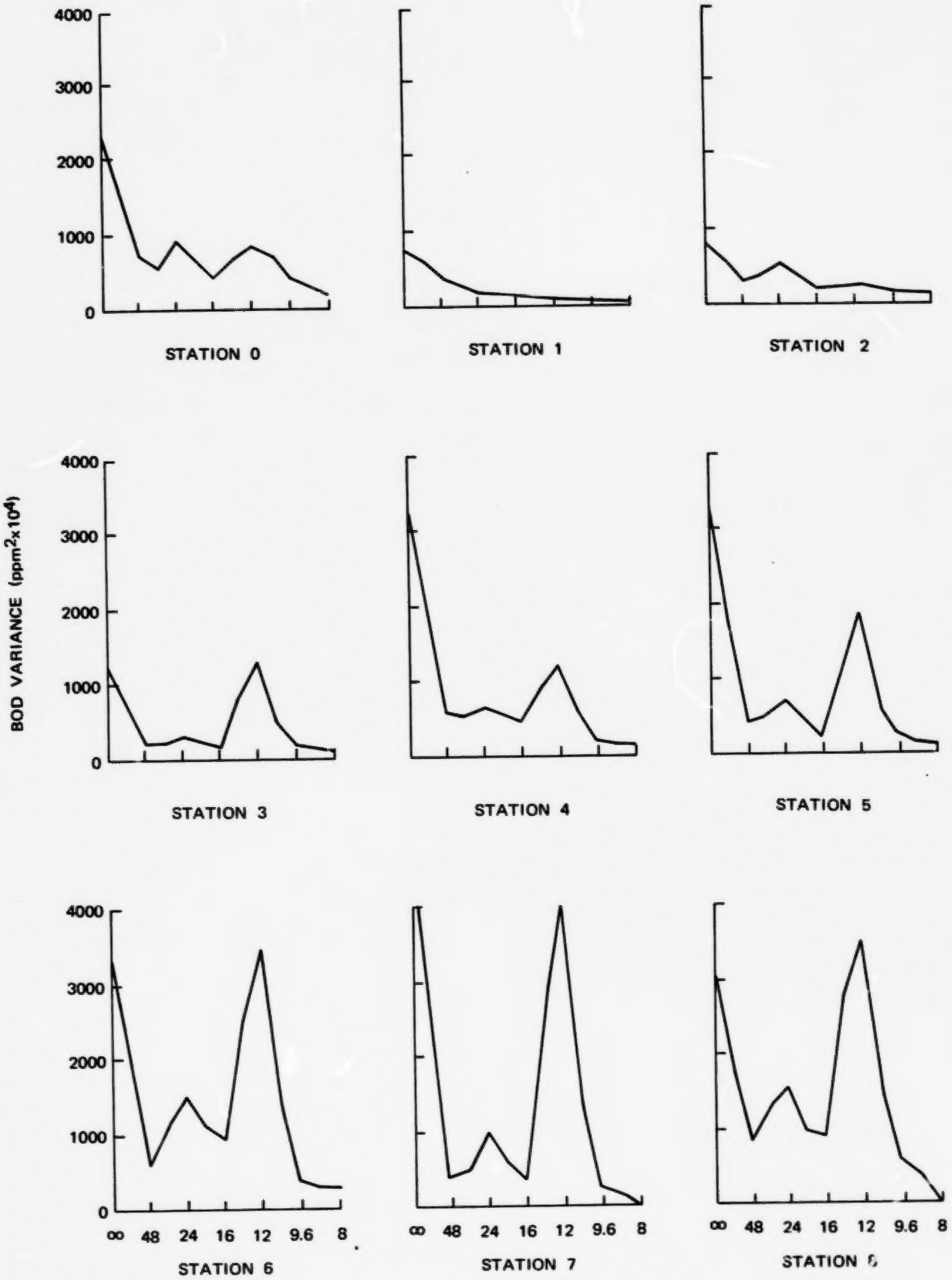


Figure 26. Stream BOD Spectra in the Potomac Estuary.

of the DO spectra themselves is also noteworthy. The diurnal variations are due to differences in algal oxygen production and consumption in the daytime and at night, while the long-period effects result from the increases and decreases in algal populations as they respond to changes in the environment.

The relationship of DO response to solar radiation with stream BOD

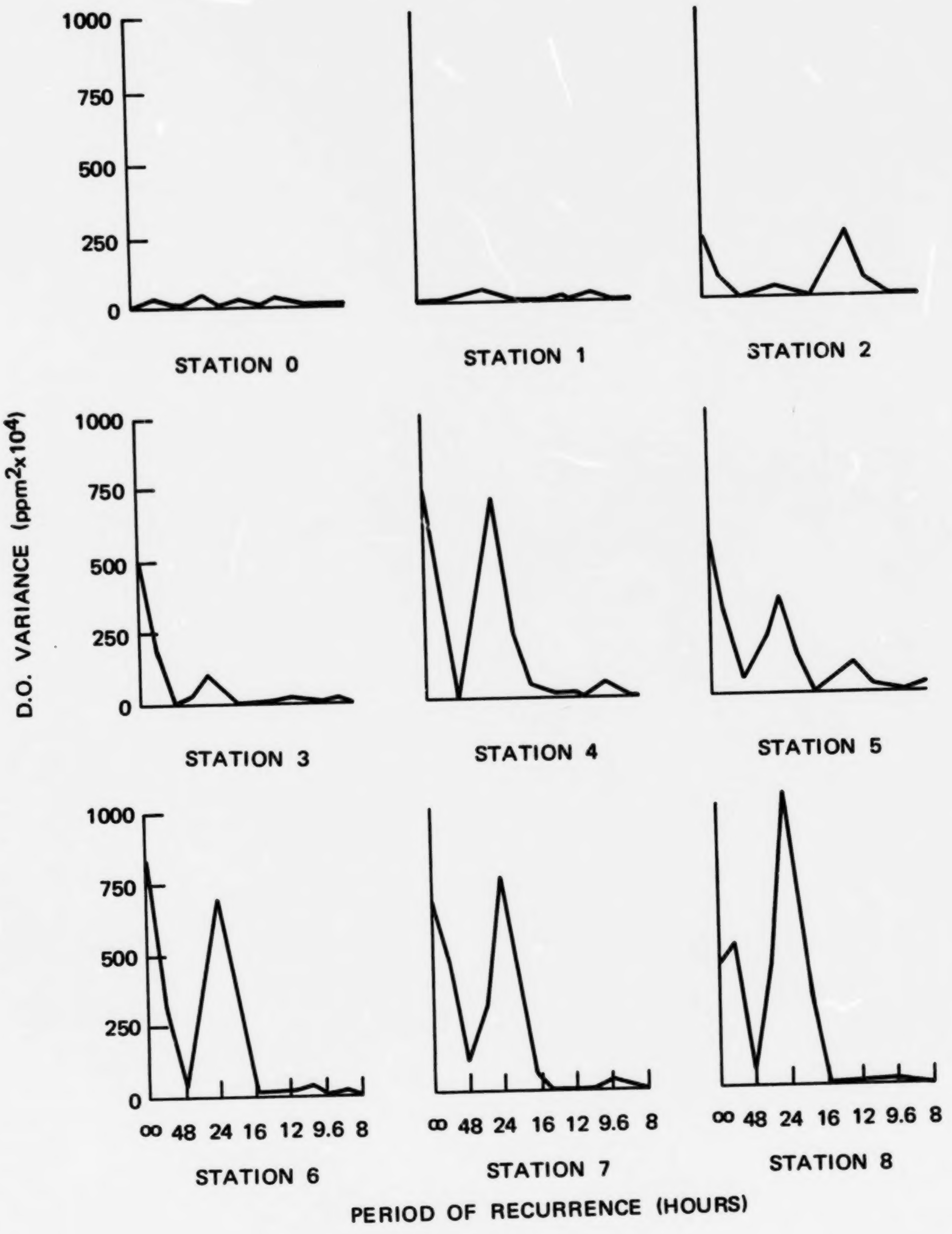


Figure 27. Stream D.O. Response to Sunlight.

may be examined by plotting the total response, *i.e.*, the sum of the responses for the individual spectral bands, at each station as a function of the mean stream BOD at each station. This has been done in Figure 28. (The point representing Station O is anomalous because of the extremely high silt turbidity of the Anascostia River.)

The sigmoid curve sketched in Figure 28 is based on the assumption that the changes in DO response are the results of changes in algal popu-

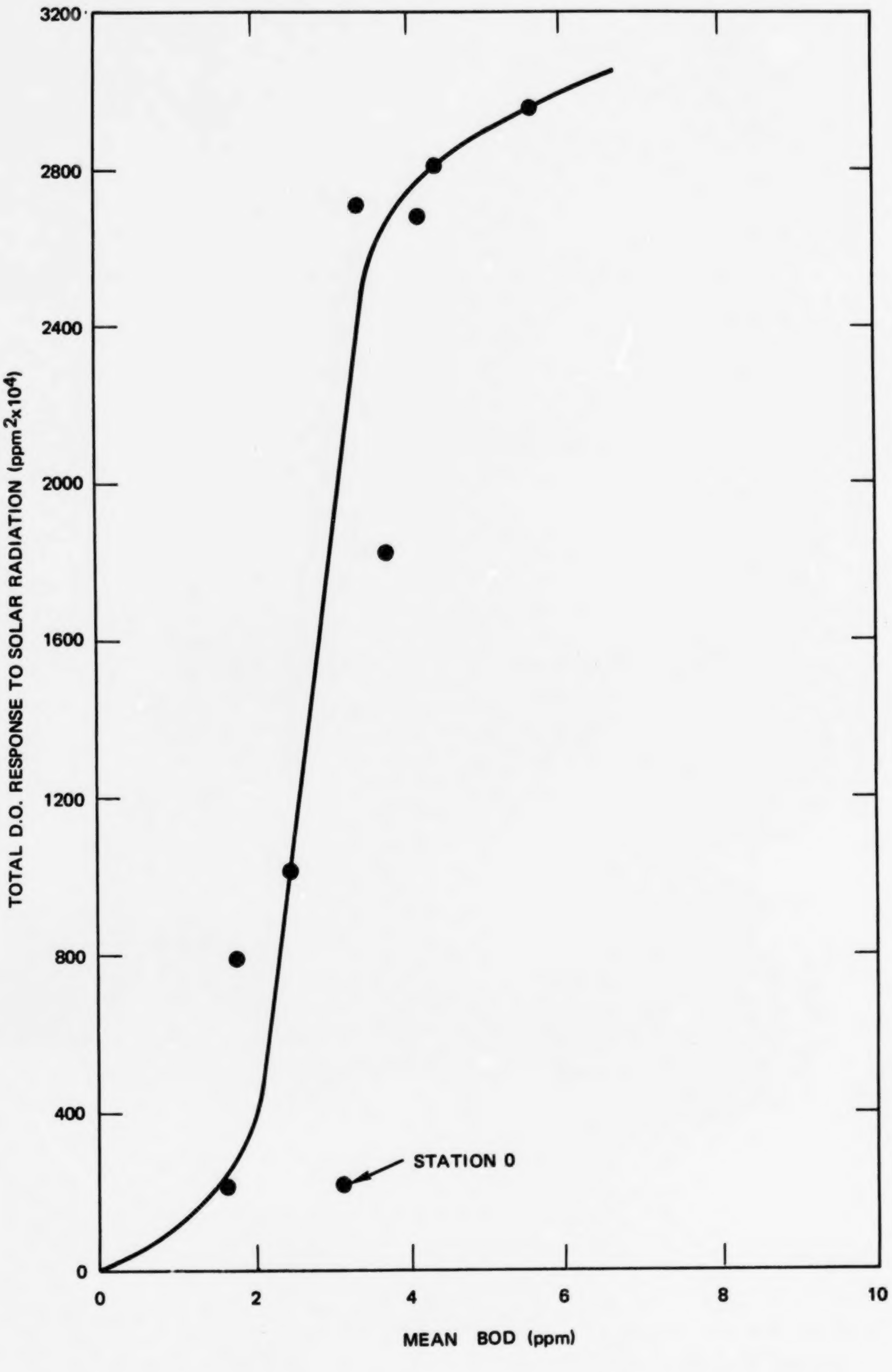


Figure 28. Relationship of BOD to D.O. Response to Solar Radiation

lation, and that the algal population is related to and is part of the stream BOD. This type of curve is generally typical of the response of a population to a favorable environment. Measurements of the algal population during this study indicated that the numbers of algae were higher at stations with higher stream BOD and that the numbers of algae fluctuated diurnally. Part of this relationship is, of course, an artifact of the BOD test.

This graph has demonstrated that the stream BOD and the solar radiation response of DO are related, and that the solar radiation is affecting the stream BOD values. (It may be safely assumed that stream BOD does not affect solar radiation.)

It would be helpful to discuss the solar radiation and stream BOD as separate driving forces on the stream DO so that quantitative estimates of the effects of waste discharges on DO can be made. To do this it is desirable to determine the extent of the the dependence of stream BOD on solar radiation and form some basis for eliminating this dependence from the total DO response to solar radiation or to BOD.

The response of BOD to sunlight was computed from the stream BOD at each station and the solar radiation. Since there is no reverse interaction of stream BOD on sunlight, we can assume that this response represents an additive effect as far as stream BOD variance is concerned, thus the effect of sunlight on stream BOD can be eliminated simply by subtracting the stream BOD response-to-sunlight spectrum from the stream BOD spectrum. This will provide a "corrected" stream BOD spectrum which can be regarded as a primary driver independent of sunlight. The net effect of solar radiation on DO is measured with the DO-solar radiation cross-spectrum; this includes the effect of the part of the stream BOD variance due to solar radiation.

The use of the DO response derived directly from the stream DO-solar radiation cross-spectra has certain great advantages over attempting to derive a DO response factor by using plankton counts, temperature, and stream BOD separately. For example, when the water temperature is highest the oxygen saturation value is lowest. This, however, might also be the time when algal oxygen production is highest, so that these two effects would tend to balance each other. While a partial correction for temperature could be made by using percent saturation of DO (at least in fresh water), this might actually obscure some relationships which could depend on the actual concentration of DO. It can be seen, therefore, that to follow the sunlight-DO system through the individual members of the chain would require a detailed knowledge of the amounts of interaction and the periods involved. While the direct computation of stream DO response to sunlight does not provide knowledge of the internal workings of this system, it does show the overall result; for many engineering applications this is all that is necessary.

The corrected stream BOD spectra are shown in Figure 29. These spectra are primarily the result of the quantities of wastes discharged into the estuary and how they mix into the receiving waters of the tidal river. The DO response spectra to corrected stream BOD can now be calculated by multiplying the DO response spectrum to BOD by $(1-H_r)$ for the solar radiation-BOD cross-spectrum. This uses the coherence as the fraction of variance related to a particular driving force. In this case the assump-

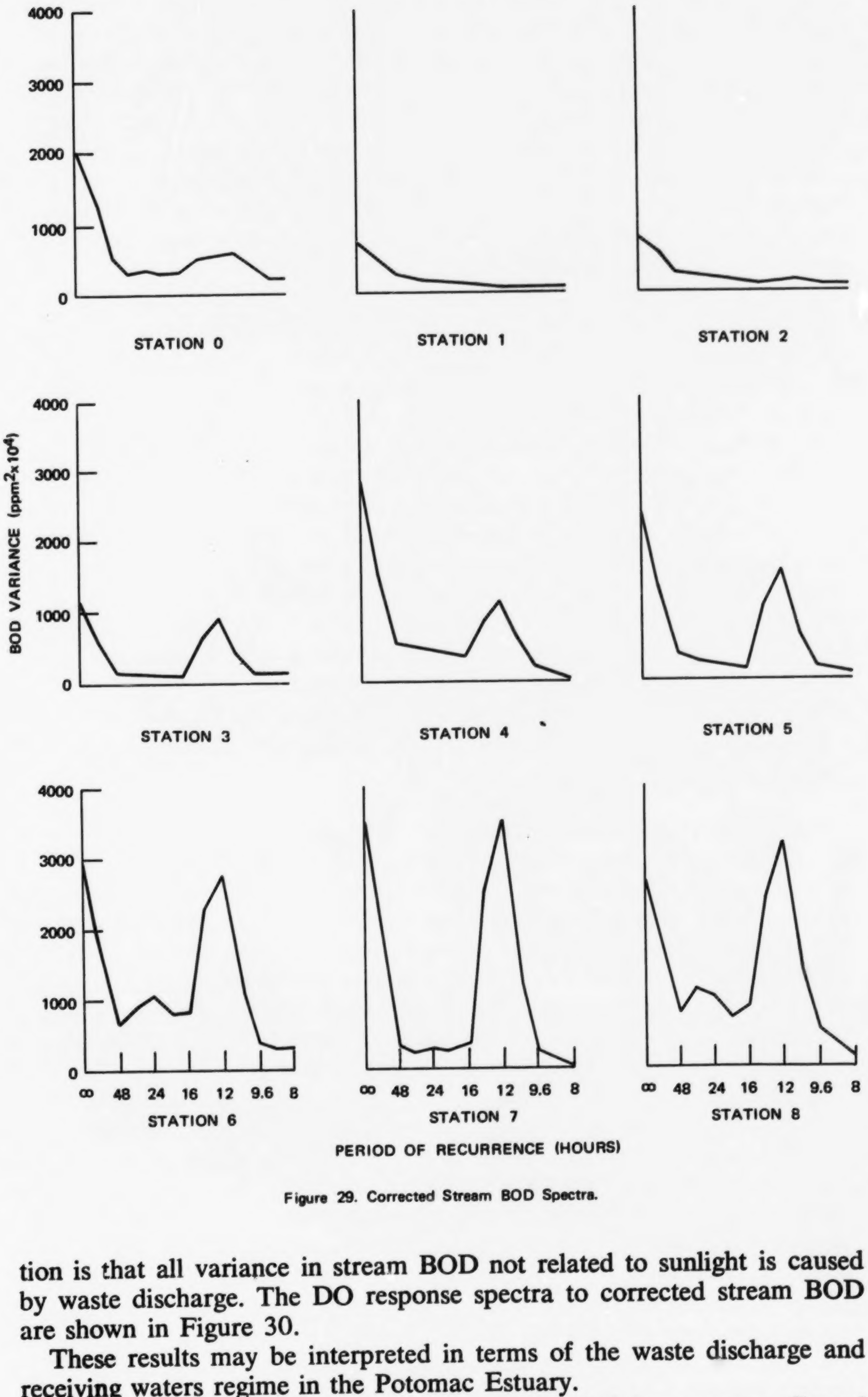


Figure 29. Corrected Stream BOD Spectra.

tion is that all variance in stream BOD not related to sunlight is caused by waste discharge. The DO response spectra to corrected stream BOD are shown in Figure 30.

These results may be interpreted in terms of the waste discharge and receiving waters regime in the Potomac Estuary.

The waste volume discharges are all dominated by a strong diurnal fluctuation, while the tidal river motion is dominated almost entirely by

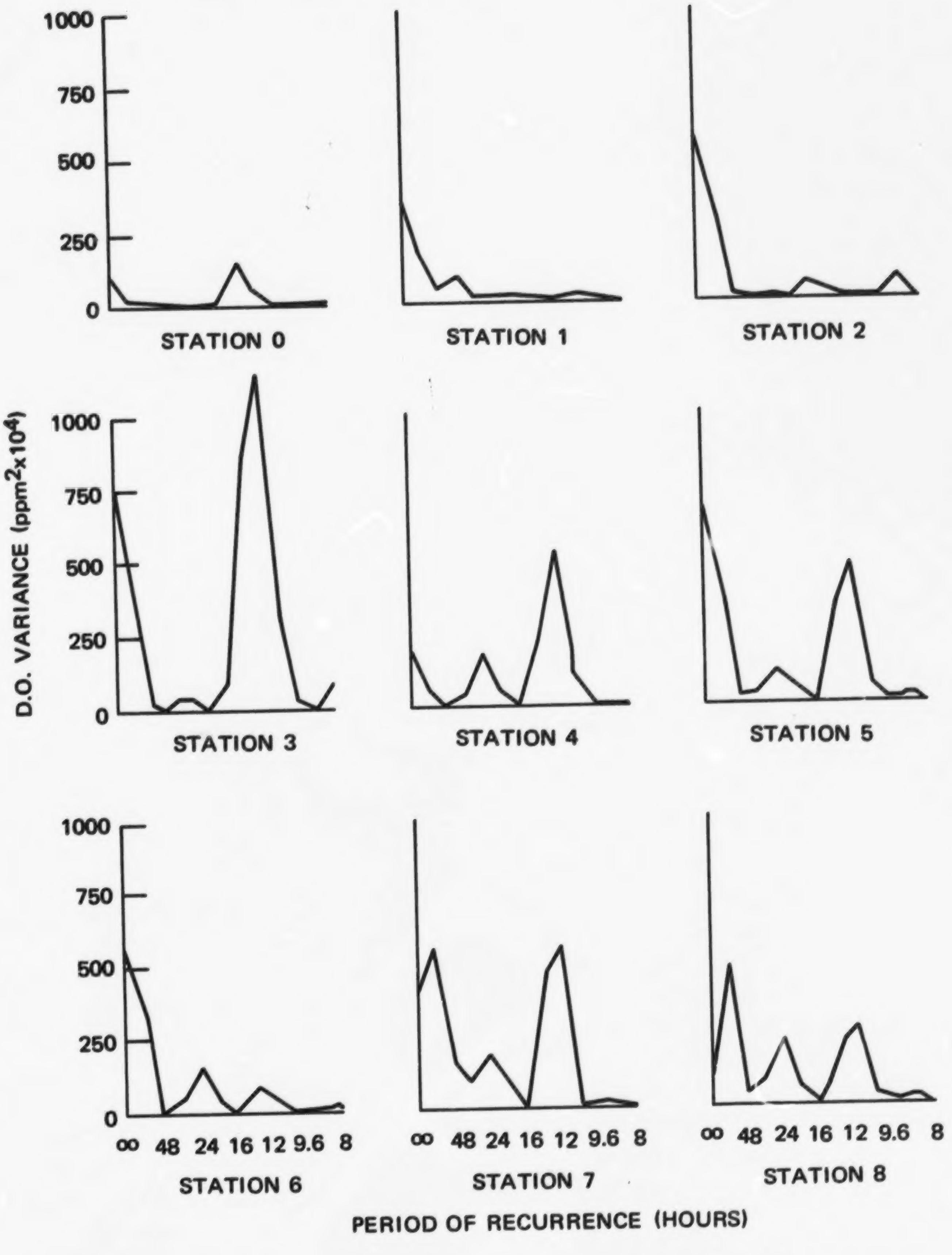


Figure 30. D.O. Response to Corrected Stream BOD

a semidiurnal rhythm. These conditions might be expected to have these effects on the stream BOD spectra: within the limits of a tidal excursion from an outfall, there may be a noticeable diurnal effect as well as a pronounced semidiurnal component. The relative magnitudes of these two components would depend upon the size relationships of the waste discharge, receiving water volume, and stream BOD already present and upon the intensity of mixing in the receiving medium. Outside the limits

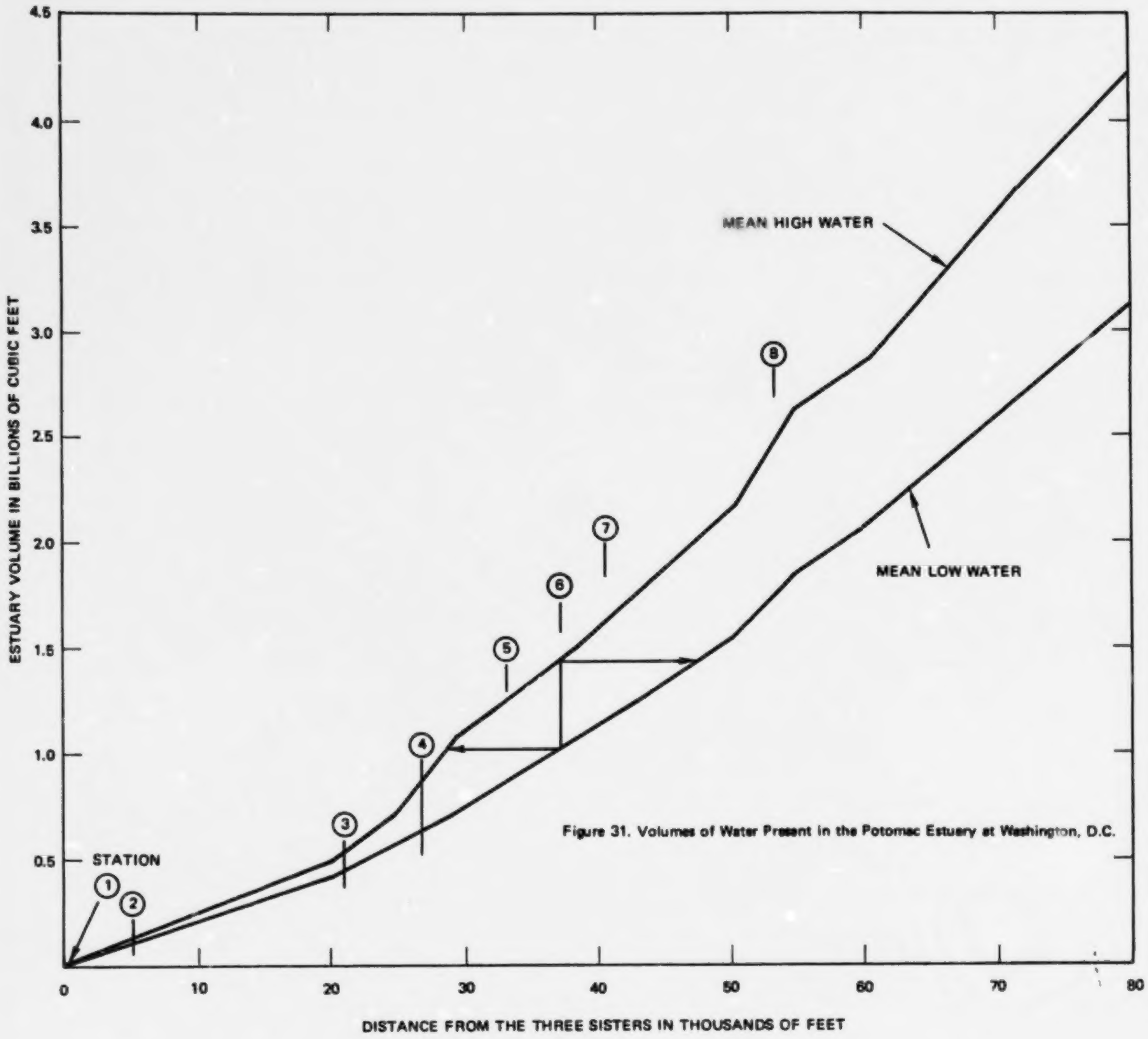


Figure 31. Volumes of Water Present in the Potomac Estuary at Washington, D.C.

of a tidal excursion from a major outfall, the stream BOD fluctuations would be dominated by the semidiurnal effect to the practical exclusion of diurnal effects. The magnitude of the semidiurnal effect observed at any station would depend on the longitudinal BOD gradient weighted by the tidal excursions; this effect might be regarded as a measure of the longitudinal turbulent mixing.

The BOD spectra for Stations 3 through 8 exhibit dominant long-period and semidiurnal effects; while the diurnal effect is noticeable at Stations 6, 8, and possibly 4, it is quite suppressed. From these spectra it might be deduced that the volume of waste discharged is quite small compared to the volume of the river into which it is discharged; the strength of the semidiurnal peaks would indicate the longitudinal BOD change is fairly small in the waters which flow past Stations 3, 4, and 5 during the tidal cycle, while the changes in the waters flowing past Stations 7 and 8 are rather large. Station 6 shows an intermediate condition. Figure 31 shows the volumes of water in the system at mean high water and at mean low water. The head of tidewater is assumed to be at Three Sisters Islands. This figure may be used to obtain a crude idea of the length of a tidal excursion in this reach of the system. For example, water opposite the Blue Plains outfall (Station 6) at high tide was about halfway between Stations 7 and 8 at low tide, while water at Station 6 at low tide lies between Stations 4 and 5 at high tide. It should be clearly understood that this type of estimate is quite crude and is based only on volume displacement. Not only are there large inherent errors in the volume measurements, but also no account is taken of the current structure of the system, the nature of the tidal wave (translatory or standing), or of the effect of river discharge. This figure is presented to give only some crude idea of what the tidal excursion might be like. It can be seen from Figure 31 that the water flowing by Stations 3, 4, and 5 during a tidal period comes primarily from the region between Station 7 and a short distance above Station 3; some water would also come from the Anacostia River. The bulk of the water involved would be represented by the mean BOD values at Stations 0, 4, 5, 6, and 7, which are 3.15, 3.34, 3.73, 4.10, and 4.30 respectively. Only a small amount of the water involved would be represented by the mean BOD of 2.44 at Station 3. It may be also noted that the BOD's from Stations 6 through 8 would begin to show the effects of larger tidal excursions, which would appear as a larger BOD change in the water moving past the station. (If the longitudinal BOD changes in the tidal excursion at each station were weighted according to the length of the tidal excursion, the results might provide some measure of the turbulent-mixing processes.)

These results and the postulated mechanism may now be considered in comparison with the DO response spectra in Figure 30. It is immediately evident that the DO response (*i.e.,* rate of change) to the diurnal BOD fluctuations is proportionately greater than to the semidiurnal fluctuations; in some cases the diurnal effects are about the same magnitude as the semidiurnal, as at Stations 6 and 8. Since the diurnal BOD component is regarded as that part of the total stream BOD which has been most recently discharged from the treatment plants, the large DO response to this component can be regarded as representative of the rapid biochemical utilization of waste materials well broken down by the treatment processes, while the gradual assimilation of the remaining materials is reflected in the long-period and semidiurnal effects.

For estimation of waste assimilation capacity it is necessary to calculate the total DO response to stream BOD. This is done by dividing the total DO response to BOD by the total corrected BOD variance at each station; the square root of this quotient is the desired factor. For the data discussed here, these results were obtained:

Station	*Response (ppm DO/ppm BOD)*
0	.22
1	.53
2	.62
3	.78
4	.39
5	.51
6	.30
7	.41
8	.29

What these factors say is that every ppm of BOD entering this section of the estuary will require about 0.5 ppm of DO. This factor is a measure of oxygen utilization by wastes in this part of the estuary and also a measure of flushing time for these wastes, if the deoxygenation rate constant is known.

These calculations with similar analyses of river discharge, tide height, waste load, and temperature records as input with the DO and BOD records as output provided the basis upon which a detailed evaluation of the effect of river discharge on the waste assimilation capacity of the Potomac Estuary could be made.

It was found that the waste assimilation capacity of the estuary increases with increasing river flow due to an increase in reaeration capability combined with a decrease in estuarine temperature which increases DO saturation values in parts of the estuary where rapid waste assimilation is taking place (*e.g.,* Stations 2 and 3).

The details of these calculations are presented in the report * from which these results were taken.

Example 4. *Relationship of Change in River Flow to Temperature and Dissolved Oxygen in an Unstratified Tidal System*

The Potomac estuary is used again to illustrate how spectral analysis provides a mechanism for estimating the magnitudes of cause and effect relationships by indirect means. In this case, with river discharge the only record available for the inflowing stream, and with temperature and DO records on the estuary but not on the river, it was possible to determine the effect of river flow on temperature and DO in the estuary, and even to estimate the DO concentration and temperature of the river above the estuary.

Temperature affects the saturation concentration value for DO in water as well as the reaction rates of some of the biological, chemical, and physical processes into which DO enters. This discussion deals primarily with the relationship of temperature to the DO saturation value. In an estuary the water temperature will be governed by two major parameters: the advection of heat into the system in inflowing water, and the absorption of thermal energy by the water from the air and from solar radiation. The DO response to temperature will then result from the action of the two primary driving forces, solar radiation and river discharge, if this is the only major source of inflowing water.

In Figure 32 the stream DO response function spectra to water temperature are plotted for each of the nine sampling stations. These spectra exhibit significant long-period, diurnal, and semidiurnal components at nearly all stations. It is useful to compare these spectra with those of the primary driving forces and of the water temperature itself. In Figure 33 are presented the spectra of solar radiation, river discharge, and tem-

---

* "Report on the Potomac River Basin Studies. Part VII, Technical Appendix." U.S. Army Corps of Engineers, 1962. (Part VII was prepared by the USPHS, R. A. Taft Sanitary Engineering Center, Cincinnati, Ohio, and Region III, USPHS, Charlottesville, Virginia.)

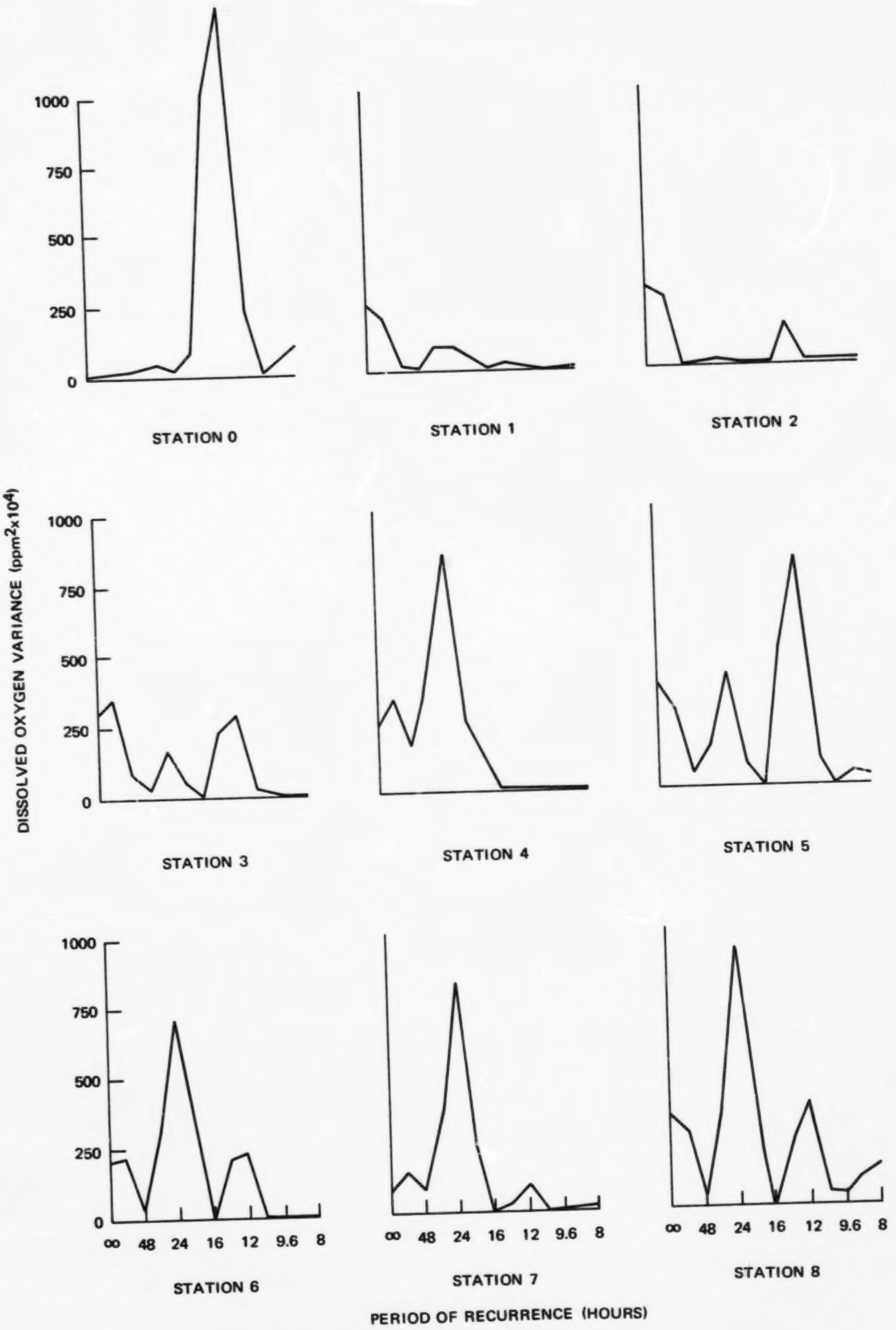
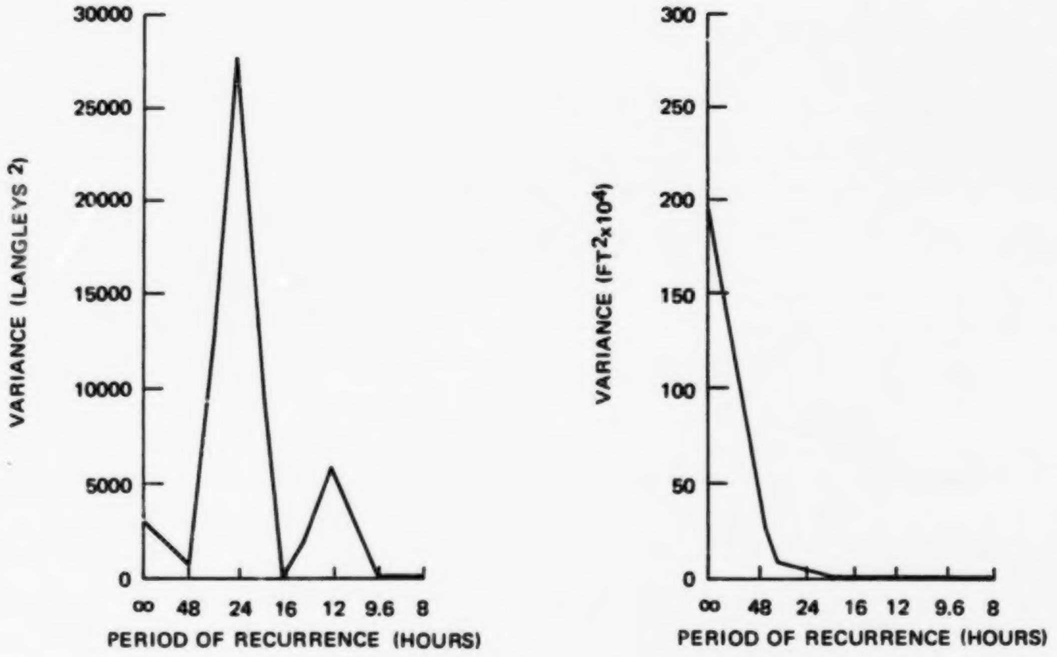
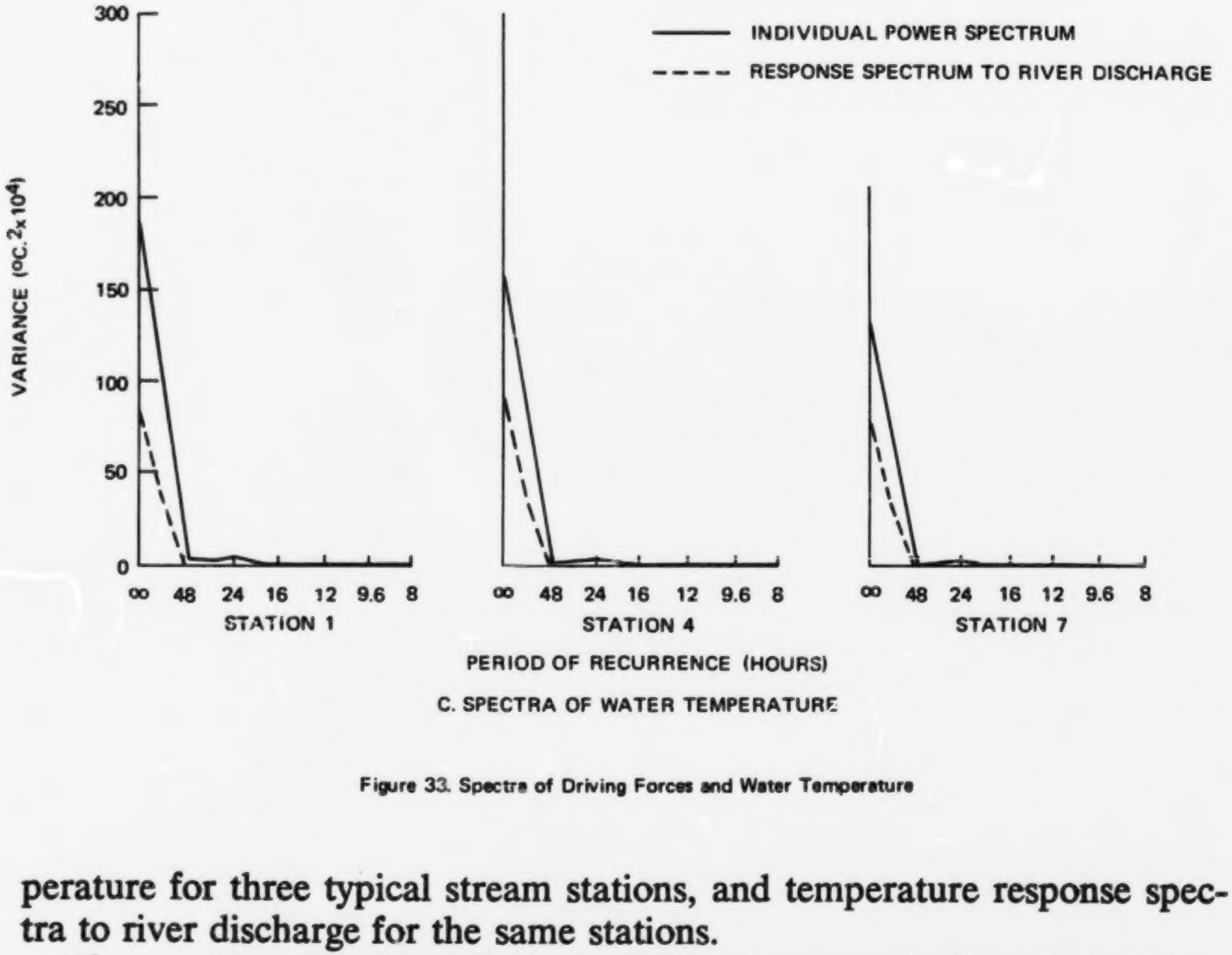


Figure 32. Dissolved Oxygen Response to Temperature



A. SPECTRUM OF SOLAR RADIATION

B. SPECTRUM OF RIVER DISCHARGE



C. SPECTRA OF WATER TEMPERATURE

Figure 33. Spectra of Driving Forces and Water Temperature

perature for three typical stream stations, and temperature response spectra to river discharge for the same stations.

The spectrum of solar radiation is primarily diurnal while that of river discharge is almost entirely long-period. The water temperature spectra are also entirely long-period, and all of the temperature response to river discharge is long-period.

From this inspection of these spectra some conclusions can be drawn about the DO response to temperature. First, it is highly unlikely that river discharge is directly related to anything but long-period response components; second, diurnal and semidiurnal components of water temperature are relatively unimportant; third, the diurnal components of the DO responses to water temperature are probably closely related to tidal flow and their magnitudes are more closely associated with the longitudinal gradients of temperature and DO than anything else.

The magnitude of the diurnal component of DO response to temperature is probably the result of a number of biological and chemical processes related to decomposition of organic matter, algal activity, and oxygen saturation values. We do not know enough about the mechanisms involved to analyze these components further. The diurnal components of the temperature spectrum are so small that a more precise determination of the diurnal change in DO due to temperature could be obtained only with much longer records than are available here.

The long-period components of the DO response spectrum lend themselves to a somewhat more detailed analysis. River discharge is the controlling parameter for these components of the DO response spectrum, so that the statistic of interest now is the DO response to river discharge. The river discharge has two effects on DO in the estuary: (1) it warms or cools the estuary, thereby changing the potential oxygen saturation values in the estuary; (2) it advects into the estuary an amount of DO which is directly proportional to its flow rate and inversely proportional to its temperature.

These effects may be examined quantitatively through cross-spectral analysis of the pertinent records. The volume of the river discharge is the only thing known about the fresh water inflow: neither the DO concentration nor the temperature is known. It is interesting, therefore, to note that the effects of river discharge on the temperature and DO of the estuary can be determined without knowing anything about the temperature and DO of the river discharge itself.

Table 9 contains the water temperature spectra and the response spectra needed for further calculations.

The phase lags are not included in any of these tables. It is noted that the temperature response to river discharge is out of phase; that is, decrease in river discharge brings about an increase in temperature. This result would be expected for August, when these data were taken. At this time of year, water resident in the estuary would absorb more heat and lose less heat to the environment than would water tumbling down from the mountains in a free flowing stream.

In Table 10 are presented some results calculated from the spectral results in Table 9.

Columns 2, 4, 5, and 7 are the long-period spectral components (long + 96 hour) taken from Tables 9.b, 9.a, 9.c, and 9.d respectively for each of the stream stations.

Column 3 is obtained from Column 2 by dividing by the variance of river discharge (0.0373), and entering 0.25 times the square root in Column 3. (The factor 0.25 is a conversion from gage height to cfs for

TABLE 9.a. *Individual Power Spectra of Water Temperature*

	T°C. × 10²								
Station	0	1	2	3	4	5	6	7	8
Period of Spectral Estimate (hours)									
Long	222	194	199	165	152	139	135	131	97.3
96	101	89.2	90.3	74.3	67.5	62.0	60.2	59.1	43.9
48	4.08	4.28	3.28	2.02	1.24	1.50	1.32	1.60	1.48
32	2.94	2.91	1.57	0.722	1.58	0.867	1.19	0.922	1.42
24	6.70	4.34	3.24	1.97	3.42	2.17	2.80	2.25	2.98
19.2	3.80	1.71	1.49	0.811	1.47	1.05	1.36	1.11	1.64
16	0.933	0.199	0.344	0.303	0.302	0.253	0.356	0.271	0.592
13.7	2.47	0.374	0.296	1.78	0.915	1.30	0.982	1.08	1.30
12	3.51	0.853	0.257	2.43	1.27	1.90	1.32	1.46	1.96
10.7	1.73	0.526	0.267	0.882	0.694	0.749	0.742	0.689	0.857
9.6	0.649	−0.0988	0.116	0.371	0.222	0.170	0.235	0.328	0.256
8.7	0.511	0.312	0.186	0.435	0.193	0.158	0.169	0.338	0.239
8	0.618	0.440	0.257	0.372	0.194	0.150	0.188	0.299	0.212
Total	351	299	301	251	231	211	206	200	154

TABLE 9.b. *Water Temperature Response to River Discharge*

	Water Temperature Response (°C. × 10^2)								
Station	0	1	2	3	4	5	6	7	8
Period of Spectral Estimate (hours)									
Long	136	84.9	91.6	91.6	86.2	78.4	73.5	73.5	50.71
96	57.0	36.4	39.3	37.0	32.9	30.4	28.3	28.5	19.75
48	0.29	0.531	0.420	0.282	0.107	0.114	0.071	0.130	0.061
32	0.165	0.180	0.078	0.071	0.099	0.030	0.0019	0.0041	0.023
24	0.836	0.466	0.661	0.209	0.397	0.198	0.335	0.265	0.363
19.2	0.302	0.0033	0.0013	0.0060	0.071	0.043	0.0019	0.0041	0.027
16	0.278	0.163	0.168	0.0102	0.021	0.011	0.0031	0.0015	0.001
13.7	0.051	0.084	0.047	0.0130	0.017	0.0012	0.0099	0.0072	0.079
12	0.032	0.223	0.236	0.427	0.074	0.129	0.051	0.040	0.004
10.7	0.178	0.115	0.019	0.0073	0.021	0.030	0.053	0.045	0.026
9.6	0.0081	0.119	0.104	0.025	0.026	0.0035	0.018	0.047	0.111
8.7	0.0758	0.025	0.0056	0.0071	0.0067	0.010	0.0064	0.0040	0.009
8	0.006	0.0076	0.0091	0.0098	0.018	0.016	0.0003	0.0033	0.002
Total	196	123	133	130	120	109	102	103	71.2
Phase Lags	Long-period effects out of phase with discharge								

TABLE 9.c. *Stream Dissolved Oxygen Response to Temperature*

(ppm^2 × 10^4)

Station	0	1	2	3	4	5	6	7	8
Period of Spectral Estimate (hours)									
Long	0.190	245	282	291	245	381	207	77.9	340
96	5.30	180	278	355	331	278	221	140.0	271
48	11.8	19.0	4.54	84.0	171	51.4	42.3	82.1	33.2
32	24.4	1.08	12.3	31.0	417	168	312	358	320
24	40.5	76.6	22.5	166	847	407	715	820	919
19.2	18.1	80.9	2.53	57.4	271	85.0	280	243	295
16	82.4	26.9	2.96	9.28	150	9.4	11.1	1.36	11.7
13.7	1010	4.98	2.65	232	13.7	538	202	34.3	217
12	1320	18.5	143	289	8.42	813	226	91.1	371
10.7	254	8.63	20.7	26.2	9.78	117	0.32	6.71	60.5
9.6	21.5	0.994	14.6	10.3	3.08	3.57	1.53	5.70	46.5
8.7	50.5	4.60	12.0	1.86	2.73	47.4	3.76	17.3	110
8	109	11.2	16.7	11.6	3.23	28.0	0.54	22.1	144
Total	2940	678	816	1560	2470	2930	2220	1900	3140

TABLE 9.d. *Stream Dissolved Oxygen Response to River Discharge*

(ppm² × 10⁴)

Station	0	1	2	3	4	5	6	7	8
Period of Spectral Estimate (hours)									
Long	0.640	340	597	194	54.1	147	109	14.6	301
96	19.7	166	327	352	488	278	305	235	329
48	10.7	70.8	38.9	32	78.4	24.9	53.5	63.2	57.6
32	0.0777	3.41	56.6	18.7	15.5	4.02	17.0	12.3	15.9
24	4.37	3.47	37.0	1.72	67.2	37.3	24.9	45.5	38.9
19.2	19.7	54.6	3.04	13.7	11.1	32.1	5.77	18.7	22.3
16	5.57	6.35	5.09	20.2	4.09	3.63	4.11	7.55	35.5
13.7	6.67	18.3	7.39	38.0	53.2	39.7	46.3	92.4	0.432
12	47.5	15.0	27.7	10.3	25.6	41.2	41.3	18.8	31.1
10.7	5.86	6.77	2.41	13.5	19.2	30.1	59.5	45.5	68.0
9.6	1.20	2.04	4.10	19.5	90.7	12.5	17.1	11.8	2.58
8.7	8.17	4.44	5.36	0.757	30.2	3.32	14.9	1.77	10.1
8	13.6	0.708	7.36	10.2	0.755	0.109	9.30	0.977	4.41
Totals	144	692	1100	724	938	654	708	568	917

TABLE 10. *Dissolved Oxygen, Temperature, and River Discharge Relationships*
In the Potomac Estuary, August 1959

	Temperature Response to River Discharge		Variance of Water Temperature	Stream DO Response to Water Temperature		Stream DO Response to River Discharge (ppm DO/1000 cfs)			
Station	(degrees C.)^2	$\frac{\text{degrees C.}}{\text{1000 cfs}}$	(degrees C.)^2	(ppm)^2	$\frac{\text{ppm}}{\text{degree C.}}$	(ppm)^2	Total	Change in DO Saturation Concentration	Change in DO due to Advection
0	1.93	1.80	3.23	.00054	.04	.00203	.058	.072	−.014
1	1.21	1.43	2.83	.0425	.124	.0506	.291	.177	.114
2	1.31	1.48	2.89	.0560	.140	.0906	.390	.207	.183
3	1.29	1.47	2.39	.0646	.165	.0546	.303	.243	.060
4	1.19	1.41	2.20	.0576	.162	.0542	.302	.228	.074
5	1.09	1.35	2.01	.0659	.181	.0425	.267	.244	.023
6	1.02	1.31	1.95	.0428	.148	.0414	.263	.194	.069
7	1.02	1.31	1.90	.0218	.107	.0250	.205	.140	.065
8	1.75	1.09	1.41	.0611	.208	.0630	.325	.227	.098
Column (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

Long-period variance of river discharge = 0.0373 the river discharge.) For Station 0 this computation is

$$\frac{1}{4}\sqrt{\frac{1.93}{0.0373}} = 1.80$$

Column 8 is obtained from Column 7 in the same way.

The entries in Column 6 are the square roots of the quotients of Column 5 divided by Column 4.

The entries in Column 9 are the products of Column 6 and Column 3.

The values presented in Column 6 have been calculated from the temperature spectra and the DO response to temperature, without use of the river discharge in the computation. It is highly significant that the mean of DO response at all the Potomac River Stations (1-8) is 0.15 ppm/degree C., the exact rate of change in DO saturation concentration given for fresh water in standard tables (*e.g., Standard Methods*). This rate of change applies at stations where the percent saturation of DO is low as well as where it is high.

The values in Column 8 are the DO responses to river discharges, which include the advected DO as well as the change due to saturation concentration of DO. By subtracting the values in Column 9 from those in Column 8, a measure of the change in DO advected to the system is obtained. Station 0 shows a decrease in DO advected with increasing discharge, probably because the greater volume of water in the main stem reduces the amount of tidal exchange in this tributary. Since the tidal prism for the estuary above any point is a fixed volume, the amount of water entering and leaving an embayment such as the Anacostia River would be inversely proportional to the river discharge entering the system in a tidal cycle.

A definite break in the magnitude of DO advected is noted between Stations 1 and 2 and the other stations on the main stem of the Potomac. The advection rates at Stations 1 and 2 indicate that the river water entering the estuary may be about 88 percent saturated with DO (based on Station 2, Column 9 divided by Column 10). An estimate of the temperature of the inflowing water may also be made. From the temperature response to river discharge at Stations 1 and 2 and the mean temperature values at these stations (28.0 and 28.1, respectively) the mean river temperature for the survey period must be about 26.6°C.; this is certainly accurate within one degree. Temperature data from a water intake at Great Falls (about 30 river miles above the estuary) showed an average water temperature for this period of 26°C. The lower advected DO values for stations farther down in the estuary reflect the fact that some of the dissolved oxygen is used up before it can get to these stations, since advective transport in an estuary is partly controlled by the reversible nature of tidal flow; a detailed analysis of these values, waste discharges, and tidal movement might, however, throw some light on the kinetics of waste assimilation in this particular system.

These results may be used to estimate the total pounds of oxygen entering the system each day in, and directly due to, the river flow. The purpose here has been merely to illustrate how the needed factors relating DO, temperature, and river discharge can be obtained by spectral analysis of the appropriate records.

Example 5. *Relationship of River Flow and Tide to Stratification in an Estuary*

This example illustrates the use of spectral analysis in attacking the classic problem in estuarine hydrodynamics; that is, changes in stratification patterns with changes in river flow and tide range.

Here also, a different approach to the use of spectral results is presented. In the first example of this chapter a rationale was presented for choosing between the *single estimate response* and the *overall response.* In the remaining examples the single estimate response and the overall response have both been used, but with an emphasis on the single estimate response approach. That is, the progression of diurnal effects was followed down the river or estuary, or the variations in semidiurnal tidal effects were considered at different points in an estuary.

This present example deals entirely with overall response values. The reason for this is that the area of concern is the transfer of energy between the long-period-dominated river discharge and the short-period-dominated tidal effects. The magnitude of this energy transfer is what will force a change in stratification pattern; the shape of the spectrum itself will result from a variety of nonlinear interactions with which present theoretical knowledge cannot cope.

The *overall response* statistic compares one total energy spectrum with another and makes it possible to avoid the complications of nonlinear interactions in solving the present problem.

The study area for this example is the harbor of Charleston, South Carolina, Figure 34; the data on which the analysis is based were obtained during six 5-day surveys at a variety of river discharges and tide ranges.* Each of the records analyzed consisted of 30 points at 4-hour sampling intervals; the analysis was run at 10 lags.

Because of the shortness of the records, an analysis of variance was run on all combinations of records to be used in the spectral analysis. The significance of the F-ratios for each interaction was used as a guide for selecting environmental parameters and water quality parameters for spectral analysis.

From the cross-spectral analysis, a coefficient for the overall response of chlorides to river discharge was computed. This coefficient indicates the magnitude of the effect of river discharge on chloride concentration and is expressed in units of gm/1/cfs. Table 11 contains the overall response coefficients calculated for each station and survey, together with the mean values of these coefficients for each survey. These coefficients include the effect of variations of all periods.

Comparison of the results from surveys AA and AB showed that the means of surface and bottom responses of chloride concentrations to river discharge were about equal for both surveys, 1.29 for AA and 1.35 for AB. On the surface the overall response was slightly higher for the spring tide condition, and on the bottom the response was slightly lower for the

---

* The complete report from which this example is taken is: "Charleston Harbor Water Quality Study." Federal Water Pollution Control Administration, USDI, June 1966.

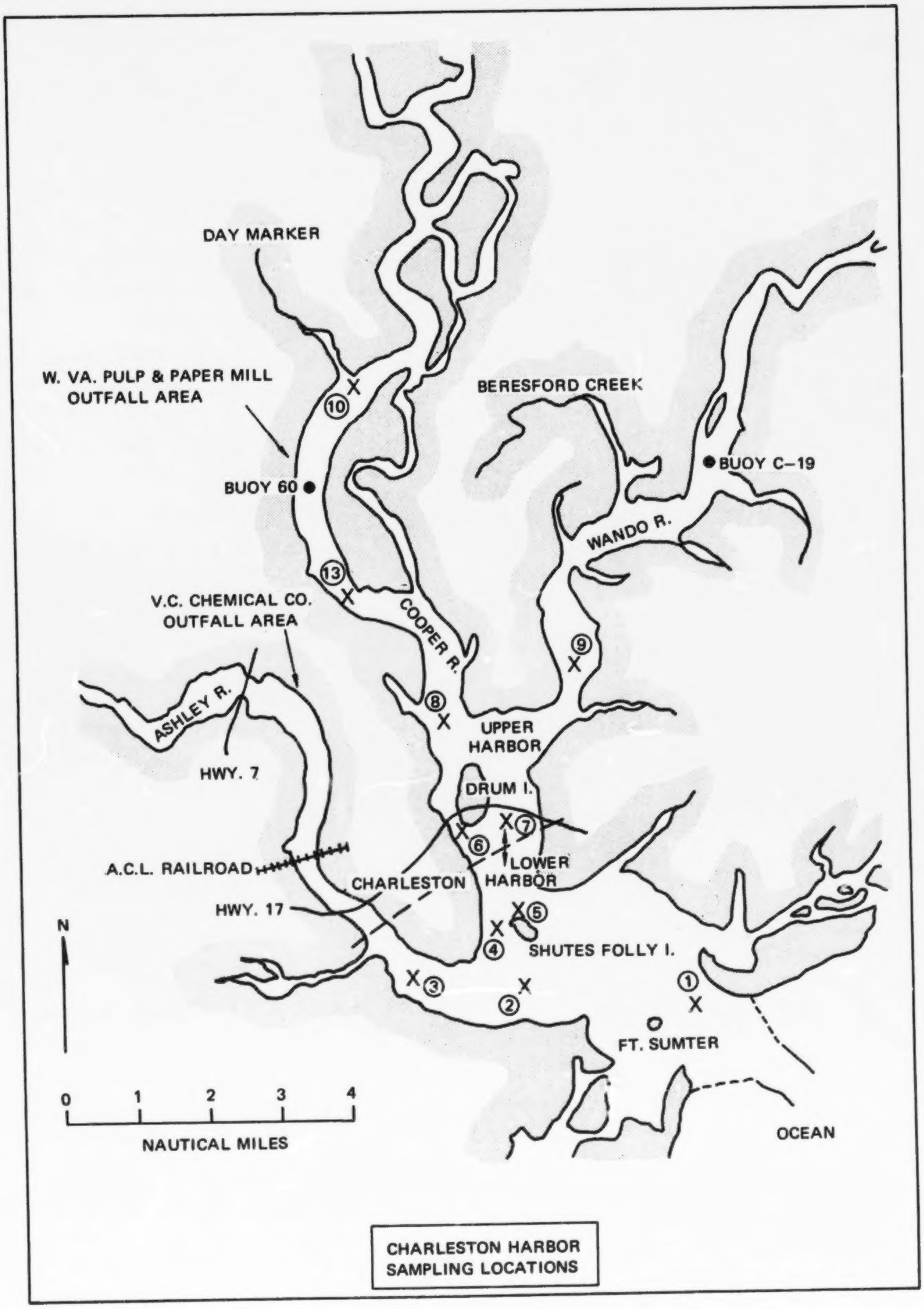

DAY MARKER
W. VA. PULP & PAPER MILL OUTFALL AREA
BERESFORD CREEK
BUOY C–19
BUOY 60
WANDO R.
V.C. CHEMICAL CO. OUTFALL AREA
COOPER R.
ASHLEY R.
HWY. 7
UPPER HARBOR
DRUM I.
A.C.L. RAILROAD
CHARLESTON
LOWER HARBOR
HWY. 17
N
SHUTES FOLLY I.
FT. SUMTER
OCEAN
0 1 2 3 4
NAUTICAL MILES
⑩ ⑬ ⑨ ⑧ ⑦ ⑥ ⑤ ④ ③ ② ①
CHARLESTON HARBOR SAMPLING LOCATIONS

Figure 34. Charleston Harbor Sampling Locations

spring tide condition. Mean surface-to-bottom response ratios of 0.555 and 0.46, respectively, showed that the estuary was well stratified during both surveys. The mean for the two surveys was 0.51. These values compared well with the measured mean chloride ratios for these surveys of 0.57 (Table 12).

TABLE 11. *Overall Response of Chloride Concentration to River Discharge, Charleston Harbor*

	(mg/1/cfs)					
Survey	AA	AB	B	C	D	E
Station						
1S	1.292	1.844	.294	.362	.103	.167
1B	1.935	1.623	.318	.298	.086	.152
2S	1.513	1.563	.115	.155	.088	.209
2B	1.475	2.139	.413	.193	.135	.203
3S	.827	.500	.088	.157	.052	.084
3B	1.638	1.852	.411	.328	.111	.133
4S	1.050	.655	.250	.156	.077	.156
4B	2.186	1.684	.300	.204	.089	.188
5S	1.637	.756	.212	.149	.090	.140
5B	2.317	1.329	.238	.254	.077	.158
6S	.737	.408	.358	.136	.060	.181
6B	1.509	1.698	.550	.245	.101	.231
7S	1.008	.758	.124	.089	.094	.166
7B	1.984	1.762	.292	.214	.137	.171
8S	.633	1.062	.165	.079	.139	.121
8B	2.451	1.931	.444	.251	.088	.157
9S	.466	.654	.192	.117	.028	.068
9B	.798	2.073	.358	.172	.053	.088
10S	.049	.354	—	—	—	—
10B	.370	2.434	—	—	—	—
13S	—	—	.030	.117	.049	.096
13B	—	—	.456	.384	.110	.206
Mean S	0.921	0.855	0.183	0.152	0.0812	0.1436
Mean B	1.666	1.853	0.378	0.254	0.0974	0.1646
Mean S & B	1.294	1.354	0.281	0.203	0.0893	0.1541
Ratio S & B	0.533	0.461	0.484	0.598	0.834	0.872

At *high* river discharges with the estuary strongly stratified the average overall response of chlorides in the estuary was dependent upon the state of the tide. During spring tide conditions there was an increase in response in the surface layer which was related to more intense mixing at the interface created by higher shearing velocities. At the same time there would have been a stronger intrusion of ocean water along the bottom, which meant less variation and consequently less response in the bottom layer.

The response spectra for the other four surveys showed generally strong semidiurnal responses with the responses at each station becoming relatively stronger on the surface as the river discharge decreased. Concurrently, the overall response of chlorides to river discharge decreased for surveys B, C, and D, and then increased during survey E (made during spring tide conditions). These results reflected a decrease in the strength of the stratification with reduced river discharge and a consequent increased mixing between surface and bottom. The main ratios of chlorides as shown in Table 12 show the same type of progression.

Surveys D and E show mean overall response *ratios* of 0.834 and

TABLE 12. *Surface- to Bottom-Chloride Ratios, Charleston Harbor*

Surveys 1965	AA 3/3-7	AB 3/23-27	B 6/21-25	C 7/19-23	D 8/16-20	E 9/20-24	Slack Tide Runs
Station 1	.707	.498	.619	.728	.720	.886	.733
Station 2	.695	.498	.529	.648	.683	.933	.820
Station 3	.754	.524	.845	.748	.789	.918	.900
Station 4	.645	.405	.045	.619	.741	.816	—
Station 5	.710	.343	.520	.721	.653	.868	.927
Station 6	.558	.212	.552	.462	.621	.808	—
Station 7	.661	.333	.497	.580	.715	.794	—
Station 8	.462	.371	.424	.417	.473	.636	—
Station 9	.838	.509	.678	.748	.830	.885	.915
Station 10	.890	.764	Discontinued		. . .	. . .	. . .
Station 13	. . .	. . .	.371	.420	.473	.645	.612
Mean Ratio of 1 thru 9	.670	.410	.572	.630	.692	.838	.784
Mean Ratios All Stations	.692	.445	.552	.609	.670	.819	—
Average River Discharge cfs	27,600	27,700	26,300	19,300	16,200	13,500	—

0.872, respectively, indicating that the surface and bottom layers reacted nearly the same to river discharge for both surveys.

The striking differences in overall *response* between surveys D and E required some examination. Table 13 shows the mean tide height range for each survey together with an analysis of the mean overall responses. The tidal height range difference between surveys D and E was about the same as that between surveys AA and AB. The surface and bottom responses for surveys AA and AB were quite different from those during surveys D and E. However, the *difference* in surface response between both pairs of surveys was practically identical, showing that the surface layer was being affected similarly by some constant difference between surveys comprising each pair. The only similarity in environmental factors was the change in tidal range.

One additional result is important; between surveys D and E the change in bottom response was the same as the change in surface response, 0.062 on the surface and 0.067 on the bottom. This comparison shows that the surface and bottom overall responses for surveys D and E were very nearly the same and that the changes of overall response with change in tide height were the same in both surface and bottom records. The *net* result of this comparison of chloride responses was to demonstrate that the surface and bottom layers in surveys D and E behaved so nearly alike that, for all practical purposes, the chloride distribution is acting as if estuary stratification was virtually nonexistent.

The cross-spectral analysis further demonstrated that, below discharges

TABLE 13. *Analysis of Chloride Responses to Tide Ranges, Charleston Harbor*

				Differences		
Survey	Mean Tide Range (ft.)	Surface Response mg/1/cfs	Bottom Response mg/1/cfs	Tide Range (ft.)	Surface Response mg/1/cfs	Bottom Response mg/1/cfs
AA	5.22	0.921	1.666	1.14	.066	0.187
AB	4.08	0.855	1.853			
B	4.09	0.183	0.378			
C	4.11	0.152	0.254			
D	4.42	0.0812	0.0974	1.34	.062	0.067
E	5.76	0.1436	0.1646			

of about 16,000 cfs, the Charleston Harbor system, considered as a single unit, behaved in its chloride distribution as if it were vertically unstratified.

For analyzing the chloride—river discharge relationship at flow rates below the range observed during the intensive survey program, the chloride *response factors* used were the average of those found during surveys D and E (0.834 and 0.872, respectively) because conditions close to vertical mixing were observed during these surveys. This would correspond to a tide range of 5.0 feet, which is near the mean tide range of 5.2 feet for the primary reference tide in Charleston Harbor.

ACKNOWLEDGEMENT

The author is deeply indebted to Dr. Blair Kinsman of the Johns Hopkins University for supervising this initiation into the mysteries of spectral analysis, and for infinite patience in observing and guiding the author's efforts at interpreting spectral results in terms of sanitary engineering phenomena.

The author is also most appreciative of the many comments and suggestions of his colleagues who have expressed interest in the use of spectral analysis; he is particularly grateful to Mr. C. G. Gunnerson for his detailed review of this material and his many thought-provoking and penetrating comments.

Some of the data presented here were obtained from field studies conducted by the USPHS and FWPCA at the request of the U.S. Army Corps of Engineers in the investigation of water resources problems in the Potomac River Basin and in Charleston Harbor, South Carolina. The permission of the Corps of Engineers to use these data is gratefully acknowledged.

BIBLIOGRAPHY

1. Anonymous. Ocean Wave Spectra. Proceedings of a conference arranged by the National Academy of Sciences. Prentice-Hall, 1963.
2. Barber, N. F. Experimental Correlograms and Fourier Transforms. Pergamon Press, New York, N.Y., 1961.
3. Bendat, J. S. Measurement and Analysis of Power Spectra and Cross-Power Spectra for Random Phenomena, M275–OU4, March 1960. Ramo-Wooldridge, Canoga Park, California.
4. Bendat, J. S. Principles and Applications of Random Noise Theory. Wiley, New York, N.Y., 1958.
5. Blackman, R. B. and J. W. Tukey. The Measurement of Power Spectra. Dover, New York, N.Y., 1958.
6. Dronkers, J. J. Tidal Computations. Interscience, 1964.
7. Goodman, N. R. On the Joint Estimation of the Spectra, Cospectrum and Quadrature Spectrum of a Two-Dimensional Stationary Gaussian Process. Scientific Paper No. 10, Engineering Statistics Laboratory, New York University, 1957. (Ph.D. Thesis, Princeton University.)
8. Grenander, U. and M. Rosenblatt. Statistical Analysis of Stationary Time Series. Wiley, New York, N.Y., 1957.
9. Hannan, E. J. Time Series Analysis. Wiley, New York, N.Y., 1960.
10. Kinsman, B. Surface Waves at Short Fetches and Low Wind Speeds—A Field Study. Chesapeake Bay Inst., Johns Hopkins University, Tech. Rep. XIX, May 1960.
11. Marks, W. and W. J. Pierson. The Power Spectrum Analysis of Ocean-Wave Records. Trans. Am. Geophys. Union 33:834–44, 1952.
12. Munk, W. H., F. E. Snodgrass, and M. J. Tucker. Spectra of Low-Frequency Ocean Waves. Bul. Scripps Inst. Oceanog. University of California 7(4):283–362, 1959.
13. Panofsky, H. A. Meteorological Applications of Power Spectrum Analysis. Bul. Am. Meteorol. Soc. 36:163–66, 1955.
14. Panofsky, H. A. and G. W. Brier. Some Applications of Statistics to Meteorology. Pennsylvania State University, 1958.
15. Press, H. and J. W. Tukey. Power Spectral Methods of Analysis and Their Application to Problems in Airplane Dynamics. Flight Test Manual, NATO, Advisory Group for Research and Development, IV–C, 1–41, June 1956.
16. Rice, S. O. Mathematical Analysis of Random Noise. Bell System Tech. 23: 282–332, July 1944; 24:46–156, January 1956. Reprinted in: Selected Paper on Noise and Stochastic Processes. N. Wax, ed. Dover, New York, N.Y., 1954.
17. Taylor, G. I. Statistical Theory of Turbulence. Proc. Roy. Soc. (London) A151:421–78, 1935.

☆ U.S. GOVERNMENT PRINTING OFFICE: O—355-390